ON THE MODELING OF MULTIPLE FUNCTIONAL OUTCOMES WITH SPATIALLY HETEROGENEOUS SHAPE CHARACTERISTICS

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Kunlaya Soiaporn

January 2014

ON THE MODELING OF MULTIPLE FUNCTIONAL OUTCOMES WITH SPATIALLY HETEROGENEOUS SHAPE CHARACTERISTICS

Kunlaya Soiaporn, Ph.D.

Cornell University 2014

This dissertation presents an approach for analyzing functional data with multiple outcomes that exhibits spatially heterogeneous shape characteristics. An example of data of this type that motivated this study is a data from a diffusion tensor imaging (DTI) study of neuronal tract in multiple sclerosis (MS) patients. DTI is an imaging technique for measuring the diffusion of water that can be used to detect abnormalities in brain tissue. DTI tractography can be summarized by 3 functional outcomes, measuring the diffusion in different directions. One of the main and most common difficulties in functional data analysis is the large number of parameters to be estimated. This is especially challenging when multiple functional outcomes are considered. To accommodate this problem, a copula approach is adopted so that the marginal distribution and the dependence structure are estimated independently. In addition to fast computation, the two-step approach also allows flexibility in the specification of the distribution of the data as the marginal distribution and copula distribution can be specified separately.

The first part of this dissertation presents an estimation algorithm using the copula approach. The marginal distribution parameters are estimated using methodology based on maximum likelihood and penalized splines. In the estimation for the dependence structure, the Karhunen-Loeve expansion and an EM algorithm are used to significantly reduce the dimension of the problem.

This allows the dependence within the same outcome and across different outcomes to be captured even in the case of many functional outcomes.

The second part of this dissertation demonstrates the application of the methodology to the DTI study. The goal is to identify the locations where the abnormalities occur and also explain the characteristics of the abnormalities in MS patients. The difference in the marginal distributions and structure dependence in the MS group from the healthy control group is then used to develop a method for predicting case status for patients.

The last part of the dissertation explores the DTI study in longitudinal setting. A larger dataset that contains DTI data from multiple visits is studied. We adopted a multilevel approach to investigate how the DTI tractography in MS patients varies over time.

BIOGRAPHICAL SKETCH

Kunlaya Soiaporn graduated from Cornell University in 2007 with bachelor's degree in Operations Research and Information Engineering. Kunlaya continued her study at Cornell OR&IE for Ph.D. in Operations Research with concentration in Applied Probability and Statistics. During Ph.D. study, besides research on modeling functional data with multiple outcomes, Kunlaya also did research on Bayesian multilevel model for ultra-high energy cosmic rays.

This document is dedicated to all Cornell graduate students.

ACKNOWLEDGEMENTS

First and foremost I would like to thank my advisor David Ruppert. I feel very privileged and grateful to be one of his Ph.D. students. Not only has he taught me much about Statistics, but he also taught me many things about professional development. Professor Ruppert has spent enormous amount of time and patience on me for the last few years while I pursued my doctorate degree. His passion and joy for quality research and perfection was contagious and motivational for me, even during my tough times. I am truly thankful for having Professor Ruppert as my Ph.D. advisor.

I would also like to express my gratitude to the co-author Raymond Carroll at Texas A&M University for offering invaluable suggestions towards my dissertation. I was fortunate to take part in the ultra-high energy cosmic ray research with Thomas Loredo, David Chernoff, and Ira Wasserman from the department of Astronomy. Working with this team and my advisor gave me invaluable experience on how to tackle a challenging problem. Their persistence in striking for perfection will always motivate me to do my best work.

During my tough and demanding graduate student life, I can always count on my OR&IE friends and my Thai friends to enjoy a small get together. May it be a weekly trip to Wegmans, or a weekend potluck, I always feel relieved of my stress with the support from my friends and colleagues.

Lastly and most importantly, I would like to express my deep gratitude to my parents and two younger sisters in Thailand. Having lived on the opposite side of the world over the past 11 years, I never once felt I was here by myself. They have always been with me regardless whether I was happy or sad. I can never thank them enough for their support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

MODELING MULTIPLE CORRELATED FUNCTIONAL OUTCOMES WITH SPATIALLY HETEROGENEOUS SHAPE CHARACTERISTICS

1.1 Introduction

Functional data analysis is a powerful tool for modeling data observed at various time points or locations to uncover the underlying features that are assumed to vary over a continuum of time points or locations. Ramsay and Silverman (2005) and Ferraty and Romain (2010) provide a comprehensive review of functional data analysis. Often we obtain multiple functional observations and we wish to study them simultaneously. In this case, it is important to have tools to study the dependence structure between different outcomes. The methodology should be efficient in reducing the dimension of the data to accommodate the possible large size of functional data, especially in the case of multiple outcomes.

This study proposes a method for studying functional data with multiple

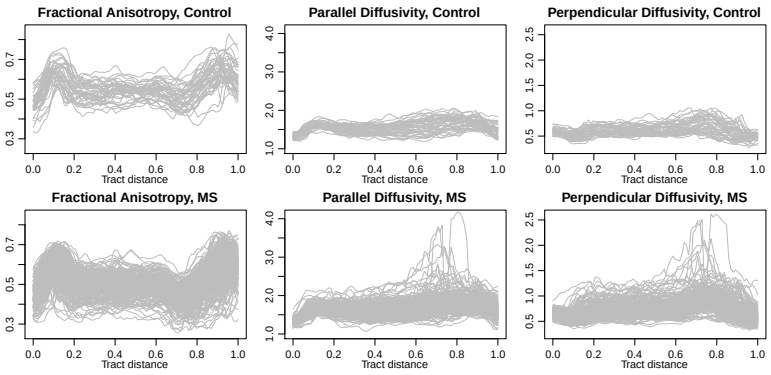

Figure 1.1: Plot of DTI data.

outcomes that exhibits some shape characteristics that vary with spatial or temporal location. Figure 1.1 illustrates an example of data that motivated this study. The data was obtained from a diffusion tensor imaging (DTI) study of neuronal tracts in 162 multiple sclerosis (MS) patients and 42 healthy controls. The top row shows the three outcomes from 42 healthy controls. The bottom row shows the same three outcomes from 162 MS patients. The tract distance is normalized to [0,1]. DTI is a magnetic resonance imaging technique for measuring the diffusion of water that can be used to detect abnormalities in brain tissue (see Basser et al., 1994, 2000). DTI tractography can be summarized by 3 functional outcomes, fractional anisotropy, parallel diffusivity and perpendicular diffusivity, each along the normalized tract distance. As seen in Figure 1.1, for each outcome, the pointwise mean and variance of both groups seem to vary along the tract location and are different between the two groups. Another interesting characteristic of this data is that the amount of skewness seems to vary along the tract location as well. An appropriate methodology to study the data should be able to capture this variation in the skewness, especially because skewness is related to disease status and could provide insights into the disease process.

Staicu et al. (2011) developed a copula-based approach for analyzing functional data with one outcome and applied their study to the parallel diffusivity measurement from this DTI data. The data were assumed to have a pointwise marginal distribution in a parametric family with shape parameters such as a skew normal or skew t family (Azzalini, 1985; Azzalini and Capitanio, 1985). The mean, standard deviation and shape parameter functions were modeled nonparametrically as functions of location. The dependence structure was estimated under the Gaussian and t copula assumption using methods based on

principal component analysis or the sample Kendall's tau matrix. Their results showed that all of the the marginal mean, variance and skewness are different between the healthy and MS groups, while the correlation within the outcome is remarkably similar between the two groups.

There are relatively few previous studies that analyze the relationship between multiple functional outcomes. Ramsay and Silverman (2005) explained and demonstrated a principal component analysis methodology to study the simultaneous variation of more than one functional outcomes. Zhou et al. (2008) used a functional data approach to analyze paired longitudinal data based on a mixed-effect model framework.

One of the main challenges in analyzing functional data with multiple outcomes is the number of parameters that need to be estimated. As the number of outcomes grows, the traditional method of moment estimators might not be practical. We overcome this challenge by using three main important tools: 1) using the copula approach introduced by Staicu et al. (2011) so that the marginal distributions of the data and its dependence structure can be modeled separately, 2) under the Gaussian copula assumption, applying the Karhunen-Loeve (KL) expansion for Gaussian processes to write the data processes as a sum of a small number of components, and then 3) using low-rank spline model for the eigenfunctions. The dependence between outcomes can be modeled through the dependence between their components. The resulting covariance structure of these components is similar to those in the study by Zhou et al. (2008) under a mixed-effect model framework. We develop a similar EM algorithm as done by Zhou et al. (2008) to calculate maximum likelihood estimators of the copula parameters. Our algorithm is fast and does not require separate steps to

model the dependence within each outcome and between different outcomes. This allows the use of a bootstrapping technique to obtain confidence intervals for the estimates. We also explain the steps needed when we want to obtain a prediction of an outcome when all the other outcomes are known, along with its confidence intervals.

The mathematical models for our methodology are presented in Section 1.2. In Section 1.3, the estimation and prediction algorithms are explained in details. Some simulation results are presented in Section 1.4 to demonstrate how our methodology performs. This chapter concludes with a discussion in Section 1.5

1.2 Modeling Multiple Outcomes

Consider a random vector $(X_1, \ldots, X_n)$, where X_i has marginal distribution $F_i(\cdot)$. This implies that for each i, $F_i(X_i)$ has a uniform distribution. A copula is a multivariate distribution function whose univariate marginal distributions are all uniform-(0,1). The copoula C of $(X_1, \ldots, X_n)$ is defined as the joint distribution of $(F_1(X_1), \ldots, F_n(X_n))$. The copula of $(X_1, \ldots, X_n)$ contains all the information about the dependence structure of $(X_1, \ldots, X_n)$. By adopting a copula approach to functional data, we can estimate the marginal distributions across the observed locations or times and the dependence structure separately. Some examples of copula are Gaussian and t- copulas. $(X_1, \ldots, X_n)$ has a Gaussian copula if $(X_1, \ldots, X_n)$ has the same copula as some Gaussian random vectors. That is $\left\{\Phi^{-1}(F_1(X_1)), \ldots, \Phi^{-1}(F_n(X_n))\right\}$ has the same distribution as some Gaussian random vector. A Gaussian copula is completely specified by a correlation matrix. Similarly $(X_1, \ldots, X_n)$ has a t-copula if $(X_1, \ldots, X_n)$ has the same copula as some

multivariate t random vector. In contrast to a Gaussian copula, a t-copula is completely specified by a correlation matrix and a degree of freedom parameter.

Let $\left\{Y_{ip}(t_{ij}); t_{ij} \in \mathcal{T}\right\}$ with $j = 1, 2, \ldots, m_i$ be the data from outcome p, $p = 1, 2, \ldots, P$ for subject i, $i = 1, 2, \ldots, N$ observed at a grid points $\{t_{i1}, \ldots, t_{im_i}\} \in \mathcal{T}$. Following the copula approach introduced by Staicu et al. (2011), we suppose that

$$Y_{ip}(t) = \mu_p(t) + \sigma_p(t) G^{-1}\left\{W_{ip}(t); \alpha_p(t)\right\} \tag{1.1}$$

where $\mu_p(t)$ is the mean function and $\sigma_p(t)$ is the standard deviation function of outcome p. Here $W_{ip}(t)$ is a latent process such that for each t, $W_{ip}(t)$ is uniformly (0,1) distributed, and $G^{-1}(\cdot, \alpha)$ is the inverse of the cumulative distribution function G in a parametric family of distribution with zero mean, unit variance and shape parameter α. For example, G can be a skew normal or skew t distribution. In the case of the skew normal, the shape parameter is the skewness parameter, while in the case of the skew t, the shape parameter includes the skewness parameter and the degree of freedom (Azzalini, 1985; Azzalini and Capitanio, 1985).

We assume that $\mu_p(t)$, $\sigma_p(t)$ and $\alpha_p(t)$ vary smoothly with t. The main objective is to estimate the marginal parameter functions and the dependence structure within each outcome and across different outcomes. We take a two-step approach. In the first step, we estimate the marginal parameter functions $\mu_p(t)$, $\sigma_p(t)$ and $\alpha_p(t)$ for each outcome p. In the second step, we estimate the correlation of $W_{ip}(t)$. The modeling of each step is explained in Sections 1.2.1 and 1.2.2, respectively.

1.2.1 Modeling Marginal Distribution

This section summarizes the method used to estimate the marginal parameter functions. For details, see Staicu et al. (2011). The pointwise marginal distributions of the data are assumed to be skew normal or skew t distribution. The estimation is done in two steps.

step 1: For simplicity, suppose the data are observed on a common dense grid of points so that $t_{ij} = t_j$ for all i and $j = 1, 2, \ldots, m$. In this first step, we obtain undersmoothed estimates $\widetilde{\mu}_p(t)$, $\widetilde{\sigma}_p(t)$ and $\widetilde{\alpha}_p(t)$ for $\mu_p(t)$, $\sigma_p(t)$ and $\alpha_p(t)$ by maximizing the pointwise likelihood function

$$\ell\left\{\mu_p(t_j), \sigma_p(t_j), \alpha_p(t_j)\right\} = \sum_{i=1}^{N} \log\left[g\left\{\frac{Y_{ip}(t_j) - \mu_p(t_j)}{\sigma_p(t_j)}\right\}; \alpha_p(t_j)\right] - \log\{\sigma_p(t_j)\} \qquad (1.2)$$

where $g(y;\alpha) = \partial G(y;\alpha)/\partial y$ is the density function corresponding to the distribution function G. When G is assumed to be a skew normal or skew-t distribution, the estimates $\widetilde{\mu}_p(t)$, $\widetilde{\sigma}_p(t)$ and $\widetilde{\alpha}_p(t)$ can be computed using the functions `sn.mle` or `st.mle` from the `R` package `sn` (Azzalini, 2011). See Staicu et al. (2011) for a discussion for the case of sparse data and alternative methods.

step 2: The estimates from step 1 are smoothed further using penalized splines. Write the mean function as $\mu_p(t) = B(t)^T\beta_{\mu,p}$, where $B(t)$ is a vector of spline functions evaluated at t and $\beta_{\mu,p}$ is a vector of spline coefficients. The estimates $\widehat{\beta}_{\mu,p}$ for $\beta_{\mu,p}$ are obtained by minimizing the penalty criterion

$$\mathrm{PL}_\mu(\beta_{\mu,p}) = \sum_{j=1}^{m}\left\{\widetilde{\mu}_p(t_j) - B_j^T\beta_{\mu,p}\right\}^2 + \lambda_{\mu,p}\Omega_{\mu,p}(\beta_{\mu,p}), \qquad (1.3)$$

where $\Omega_{\mu,p}(\beta_{\mu,p}) = \beta_{\mu,p}^T D_{\mu,p}\beta_{\mu,t}$, $B_j = B(t_j)$, and $D_{\mu,p}$ is a penalty matrix. See Ruppert et al. (2003) for a discussion of penalized splines and penalty matrices. The final estimates $\widehat{\mu}_p$ is obtained as $B(t)^T\widehat{\beta}_{\mu,p}$. The estimates for the standard deviation function can be computed similarly. For the shape parameter function,

Staicu et al. (2011) proposed a penalized marginal pseudo-likelihood criterion where the mean and variance parameter functions are fixed at estimates. That is, the criterion to be minimized are

$$\mathrm{PL}_\alpha(\beta_{\alpha,p}) = -2\sum_{i=1}^{N}\sum_{j=1}^{m}[\ell_{ij}\{\beta_{\alpha,p}; \widehat{Y}_{ip}(t_j)\}] + \lambda_{\alpha,p}\Omega_{\alpha,p}(\beta_{\alpha,p}) \tag{1.4}$$

where $\widehat{Y}_{ip}(t_j) = \left\{Y_{ip}(t_j) - \widehat{\mu}_p(t_j\right\}/\widehat{\sigma}_p(t_j)$ are the standardized observations and $\ell_{ij}(\beta_{\alpha,p}; \widehat{Y}_{ip}(t_j)) = \log[g\{\widehat{Y}_{ip}(t_j); B_j^T\beta_{\alpha,p}\}]$. In some cases, it might be better to model a transformed parameter $h(\alpha)$ instead of α. Some possible criteria for choosing smoothing parameters are the restricted maximum likelihood (REML) (Wood, 2006), AIC, corrected AIC (Ruppert et al., 2003) and cross-validation (CV). See Staicu et al. (2011) for details and a discussion of smoothing parameter selection and an alternative method.

1.2.2 Modeling the Dependence across Different Outcomes

We assume Gaussian copulas for simplicity and because they seem adequate for our purposes. After the estimates of the parameters of marginal distributions for each outcome $\widehat{\mu}_p(t)$, $\widehat{\sigma}_p(t)$ and $\widehat{\alpha}_p(t)$ are obtained, we transform the observed outcome by

$$R_{ip}(t) = \Phi^{-1}\left[G\left\{\frac{Y_{ip}(t) - \hat{\mu}_p(t)}{\hat{\sigma}_p(t)}; \hat{\alpha}_p(t)\right\}\right], \tag{1.5}$$

where Φ is the cumulative distribution function of the standard normal distribution, and $G(\cdot;\alpha)$ is the distribution function of a distribution with mean 0, variance 1, and shape parameter α. We model R_{ip} as the sum of two independent components: 1) a finite Karhunen-Loeve (KL) expansion with a small number

of components, and 2) a white noise process with variance $\sigma^2_{\epsilon p}$. That is,

$$R_{ip}(t) = \sum_{k=1}^{K_p} Z_{ipk} f_{kp}(t) + \epsilon_{ip}(t), \tag{1.6}$$

where $\{Z_{ikp}, k = 1, 2, \ldots, K_p\}$ are normally distributed with mean 0 and are independent across k, and f_{kp} are the eigenfunctions of the covariance function of the process $R_{ip}(t)$ (Levy, 2008). Since the marginal variance of $R_{ip}(t)$ equal to 1, it follows that for all t,

$$\sum_{k=1}^{K_p} f^2_{kp}(t)\text{var}(Z_{ipk}) + \sigma^2_{\epsilon p} = 1. \tag{1.7}$$

For identifiability purpose, we assume that the variances of $\{Z_{ikp}, k = 1, 2, \ldots, K_p\}$ are in decreasing order.

The eigenfunctions f_{kp} can be estimated using splines as follows. Let $b(t) = \{b_1(t), \ldots, b_q(t)\}^T$ be an orthonormal spline basis, i.e. $\int_{\mathcal{T}} b_k(t) b_l(t) dt = \delta_{kl}$. We estimate $f_p = (f_{1p}, \ldots, f_{K_p,p})^T$ using $b(t)$ as

$$f_p(t)^T = b(t)^T \Theta_p \tag{1.8}$$

where Θ_p is a matrix of spline coefficients with dimension $q \times K_p$ with orthogonal columns. The orthogonality of $b(t)$ and Θ_p implies the orthogonality of the principal component curves f_{kp}.

Let $Z_{ip} = \left(Z_{ip1}, \ldots, Z_{ipK_p}\right)^T$ and $Z_i = \left(Z^T_{i1}, \ldots, Z^T_{iP}\right)^T$. Let $D_p = \text{cov}(Z_{ip})$, which is diagonal since Z_{ipk} are independent across k. To have identifiability, we also have that the diagonal elements of D_p are in a decreasing order. Let $C_{pp'} = \text{cov}(Z_{ip}, Z_{ip'})$. Then we can write the covariance matrix Σ of Z_i as

$$\Sigma = \text{cov}(Z_i) = \begin{pmatrix} D_1 & C_{12} & \cdots & C_{1P} \\ C_{21} & D_2 & \cdots & C_{2P} \\ \vdots & \vdots & \ddots & \vdots \\ C_{P1} & C_{P2} & \cdots & D_P \end{pmatrix} \tag{1.9}$$

Suppose that we have the same observation times $\{t_j, j = 1, 2, \ldots, m\}$ for every subject i. Let $R_{ip} = \left\{R_{ip}(t_1), \ldots, R_{ip}(t_m)\right\}^T$, $B = \{b(t_1), \ldots, b(t_m)\}^T$ and $\epsilon_{ip} = \left\{\epsilon_{ip}(t_1), \ldots, \epsilon_{ip}(t_m)\right\}^T$. Putting everything together, we have the following reduced-rank model for R_{ip},

$$\begin{aligned} R_{ip} &= B\Theta_p Z_{ip} + \epsilon_{ip} \qquad (1.10) \\ \epsilon_{ip} \sim \mathrm{N}(0, \sigma^2_{\epsilon p} I_m), \quad Z_{ip} &\sim \mathrm{N}(0, D_p), \quad \mathrm{cov}(Z_{ip}, Z_{ip'}) = C_{pp'}, \quad \text{for} \quad p \neq p'. \end{aligned}$$

Our reduced-rank model for the latent processes (1.10) has a similar dependence structure as the model for two outcomes introduced by Zhou et al. (2008). Here we adopt a similar approach using an iterative EM algorithm for estimating the parameters.

To have identifiability, it is sufficient to require that the first nonzero elements of each column of Θ_p are positive. With finite samples, it is best to determine the sign using the elements with the largest magnitude in each column of Θ_p, since this choice is least influenced by finite-sample random fluctuation (for details, see Zhou et al., 2008). In our estimation algorithm, we require that, in every iteration, the element with the largest magnitude in each column of Θ_p is positive. In the next section, we develop an algorithm for estimating parameters in model (1.10).

1.3 Estimation Algorithms

This section explains the criterion and algorithm for estimating the parameters for copula , along with a discussion of tuning parameter specification. We also explain how to predict an outcome after we have observed each of the other

outcomes.

1.3.1 Parameter Estimation Algorithm

The estimates for the dependence parameters can be obtained by maximizing the pseudo-likelihood, with R_{ip} obtained by transforming the observed data using the estimates of the marginal parameters. Let $L_i(\Lambda)$ denote the contribution to the likelihood from individual i, where Λ is the covariance matrix of $R_i = (R_{i1}^T, \ldots, R_{iP}^T)^T$. The loglikelihood for individual i is given as

$$\log L_i(\Lambda) = -\frac{Pm}{2}\log(2\pi) - \frac{1}{2}\log|\Lambda| - \frac{1}{2}R_i^T\Lambda^{-1}R_i, \tag{1.11}$$

where Λ can be computed in terms of $\left(\left\{\Theta_p\right\}, \Sigma, \left\{\sigma_{\epsilon p}^2\right\}\right)$ as

$$\operatorname{cov}(R_i) = \Lambda = \begin{pmatrix} \Lambda_{11} & \Lambda_{12} & \cdots & \Lambda_{1P} \\ \Lambda_{21} & \Lambda_{22} & \cdots & \Lambda_{2P} \\ \vdots & \cdots & \ddots & \vdots \\ \Lambda_{P1} & \Lambda_{P2} & \cdots & \Lambda_{PP} \end{pmatrix} \tag{1.12}$$

where

$$\Lambda_{pp} = B\Theta_p D_p \Theta_p^T B^T + \sigma_{\epsilon p}^2 I_m \qquad \text{and} \qquad \Lambda_{pp'} = B\Theta_p C_{pp'} \Theta_{p'}^T B^T \tag{1.13}$$

Even though the main interest here is to estimate the covariance matrix Λ of R_i, not the principal component functions which can obtained from Θ_p, we use the reduced-rank model for R_{ip} so that our estimation algorithm involves estimating D_p, $C_{pp'}$ and Θ_p which are of much smaller dimension than Λ. The KL expansion also provides a nice diagonal structure for D_p. Our iterative algorithm estimates D_p and $C_{pp'}$ sequentially. Once the estimates for all of the parameters are obtained, the estimated covariance of $R_{ip}(t)$ can be computed by plugging in the estimates to equation (1.13).

Extending the approach as in Zhou et al. (2008), we use the penalized pseudo-likelihood and minimize

$$\begin{aligned} & -2\sum_{i=1}^{N}\log L_i(\Lambda) + \sum_{p=1}^{P}\lambda_p\sum_{k=1}^{K_p}\Theta_{pk}^T\int b''(t)b''(t)^T dt\Theta_{pk} \\ = & \sum_{i=1}^{N}\left\{Pm\log(2\pi) + \log|\Lambda| + R_i^T\Lambda^{-1}R_i\right\} + \sum_{p=1}^{P}\sum_{k=1}^{K_p}\lambda_{pk}\Theta_{pk}^T\int b''(t)b''(t)^T dt\Theta_{pk}, \quad (1.14) \end{aligned}$$

where Θ_{pk} denotes the k^{th} column of Θ_p, and λ_{pk} are the smoothing parameters. In general, we can have different λ_{pk} for each f_{pk}. Since smoothing parameters are mostly determined by the scale of data, to simplify the computation, we assume that $\lambda_{pk} = \lambda_p$ for all $k = 1, \ldots, K_p$. Minimizing this expression can be complicated. Instead, we treat the Z_i as missing values and use the EM algorithm (Dempster et al., 1977). The joint loglikelihood is given as

$$\begin{aligned} \log L(R_i, Z_i) &= \log\left\{f(R_{i1}|Z_{i1})\ldots f(R_{ip}|Z_{ip})f(Z_i)\right\} \\ &= \sum_{p=1}^{P}\left\{\frac{-m}{2}\log(\sigma_{\epsilon p}^2) - \frac{1}{2\sigma_{\epsilon p}^2}(R_{ip} - B\Theta_p Z_{ip})^T(R_{ip} - B\Theta_p Z_{ip})\right\} \\ &\quad -\frac{1}{2}\log(|\Sigma|) - \frac{1}{2}Z_i^T\Sigma^{-1}Z_i. \quad (1.15) \end{aligned}$$

The EM algorithm is performed as follows:

The E-step: Compute the conditional distribution of Z_i given R_i

The M-Step: Update the parameter estimates by minimizing

$$-2E\left\{\sum_{i=1}^{N}\log L(R_i, Z_i)\Big|R_i\right\} + \sum_{p=1}^{P}\lambda_p\sum_{k=1}^{K_p}\Theta_{pk}^T\int b''(t)b''(t)^T dt\Theta_{pk}. \quad (1.16)$$

The details for each step are given in 3.4.

1.3.2 Specification of Splines and Tuning Parameters

Recall that we require the splines $b(t)$ used to estimate the eigenfunctions $f_p(t)$ to be orthonormal; see equation (1.8). This implies that the matrix $B = \{b(t_1), \ldots, b(t_m)\}^T$ has to be orthogornal. As in Zhou et al. (2008), to obtain an orthogonal matrix B, we start by choosing any spline basis $\widetilde{b}(t)$, such as a truncated power basis, then we evaluate these functions at the observation points to obtain the matrix $\widetilde{B} = \{\widetilde{b}(t_1), \ldots, \widetilde{b}(t_m)\}^T$. Let $\widetilde{B} = QR$ be the QR decomposition of $\widetilde{B}$, where Q is an orthogonal matrix and R is an upper triangular matrix. Then the orthogonal matrix B is obtained by $B = \widetilde{B}R^{-1}$. `R` package `orthogonalsplinebasis` (Redd, 2011) provides functions for orthogonalizing a spline basis.

The two sets of tuning parameters that need to be specified are the spline smoothing parameters, λ_p, and the number of components in the KL expansion, K_p. For penalized splines, the knots are typically placed at fixed quantiles of the observation times or locations, and the number of knots are not critical provided that it is sufficient large. The smoothness of the curves is mainly determined by the smoothing parameters λ_p (Ruppert, 2002; Ruppert et al., 2003). We start by performing the analysis for each individual outcome. For each outcome p, we specify a range for λ_p to perform grid search. For each λ_p fixed at a value in the grid search range, we vary the number of components starting by using only one component. The EM algorithm (as explained in Section 1.3.1) is used to fit each 1-outcome curve. We keep adding another component and performing the EM algorithm until the variance of the newly added principal component score is less than a prespecified small fraction of that of the previously added component. The number of components before adding this last component is K_p

chosen for this λ_p. The corrected AIC (Ruppert et al., 2003) is computed for this (λ_p, K_p). The pair (λ_p, K_p) with the minimum corrected AIC is selected for this outcome. Other criteria such as CV and AIC can also be used. Once we analyze multiple outcomes, the number of components can be fine tuned further. We do not expect the values of λ to change from the individual outcome case when we do multiple outcome analysis, as λ should depend on the scale of individual outcome data more than the dependence structure between outcomes.

1.3.3 Confidence Intervals for Estimates

Bootstrapping can be easily applied to obtain pointwise confidence intervals for the estimates for the marginal distribution parameters and the correlations. Bootstrap samples are obtained by resampling the subjects from the original data $\left\{Y_{ip}\right\}$. For each bootstrap sample, we calculate estimates for the marginal distribution parameters (as explained in section 1.2.1), use them to transform the data as in equation (1.5), and then perform the EM algorithm to obtain estimates for the correlations, by plugging in the parameter estimates to equation (1.13). The pointwise confidence intervals for the estimates of both the marginal distribution parameters and the correlations are computed based on sample quantiles of the estimates for the bootstrap samples.

1.3.4 Prediction of an Outcome Using all the Other Outcomes

In addition to simplifying the computation for parameter estimation, our reduced-rank model for multiple outcomes also provides a straightforward way

of predicting an outcome for an individual when all the other outcomes of the same subject are known. This can be done using the conditional distribution property of the multivariate normal random variables as follows. Suppose we know the observation Y_{ip} of a subject i for all outcomes $p \neq 1$ and we want to predict outcome 1 for this individual. Let R_{ip} denote the transformed process as in equation (1.5). Since we have that $(R_{i1}, \ldots, R_{iP}) \sim \mathrm{N}(0, \Lambda)$, we can compute the conditional distribution of R_{i1} as

$$R_{i1}|R_{i2}, \ldots, R_{iP} \sim \mathrm{N}\left(\bar{\mu}_1, \bar{\Sigma}_1\right) \tag{1.17}$$

where

$$\bar{\mu}_1 = \begin{pmatrix} \Lambda_{12} & \ldots & \Lambda_{1P} \end{pmatrix} \begin{pmatrix} \Lambda_{22} & \ldots & \Lambda_{2P} \\ \vdots & \ddots & \vdots \\ \Lambda_{2P} & \ldots & \Lambda_{PP} \end{pmatrix}^{-1} \begin{pmatrix} R_{i2} \\ \vdots \\ R_{iP} \end{pmatrix}, \tag{1.18}$$

and

$$\bar{\Sigma}_1 = \Lambda_{11} - \begin{pmatrix} \Lambda_{12} & \ldots & \Lambda_{1P} \end{pmatrix} \begin{pmatrix} \Lambda_{22} & \ldots & \Lambda_{2P} \\ \vdots & \ddots & \vdots \\ \Lambda_{2P} & \ldots & \Lambda_{PP} \end{pmatrix}^{-1} \begin{pmatrix} \Lambda_{21} \\ \vdots \\ \Lambda_{P1} \end{pmatrix}. \tag{1.19}$$

To obtain a prediction $\widehat{Y}_{i1}$ for Y_{i1}, generate $R_{i1}^{(s)}$, for $s = 1, 2, \ldots, S$, from $\mathrm{N}\left(\bar{\mu}_1, \bar{\Sigma}_1\right)$ distribution, where S is sufficiently large. The predicted value is computed as

$$\widehat{Y}_{i1}(t) = \frac{1}{S}\left[\sum_{s=1}^{S} G^{-1}\left\{\Phi\left(R_{i1}^{(s)}\right); \widehat{\alpha}_1(t)\right\}\right]\widehat{\sigma}_1(t) + \widehat{\mu}_1(t) \tag{1.20}$$

The variance of prediction can be computed using bootstrapping. The prediction for other outcomes can be done similarly.

1.4 Simulation Studies

In this section, we demonstrate our methodology using simulated data. As far as we are aware, there exist no other estimators for multiple functional outcomes of the type we are considering, that is, with spatially varying non-Gaussian characteristics. Therefore, there are no competing estimators to compare with ours.

The simulated data contains 100 datasets, each containing observations of $P = 3$ outcomes from $N = 200$ subjects. Each observation is taken at the common time points t_j, $j = 1, 2, \ldots, 80$. All of the data are generated using the finite KL expansion for Gaussian processes as follows:

1. Generate

$$Q_{ip}(t) = \sum_{k=1}^{K_p} Z_{ipk} h_{kp}(t) + \epsilon_{ip}(t) \tag{1.21}$$

for $p = 1, 2, 3$, where $K_1 = K_2 = 2, K_3 = 3, \epsilon_{ip}(t) \sim N(0, \tau_p^2), \tau_1^2 = 0.2, \tau_2^2 = 0.3, \tau_3^2 = 0.4, Z_i = (Z_{i11}, Z_{i12}, Z_{i21}, Z_{i22}, Z_{i31}, Z_{i32}, Z_{i33})$ is normal with mean 0 and covariance matrix

$$\left(\begin{array}{cc|cc|ccc}
6.47 & 0 & 4.46 & 1.53 & 2.93 & 0.31 & 1.76 \\
0 & 4.08 & 0.02 & 0.89 & 0.99 & -2.39 & -1.14 \\
\hline
-4.46 & 0.02 & 20.03 & 0 & -3.52 & 2.61 & -4.05 \\
1.53 & 0.89 & 0 & 7.46 & 3.51 & -2.01 & -1.43 \\
\hline
2.93 & 0.99 & -3.52 & 3.51 & 16.34 & 0 & 0 \\
0.31 & -2.39 & 2.61 & -2.01 & 0 & 7.63 & 0 \\
1.76 & -1.14 & -4.05 & -1.43 & 0 & 0 & 2.98
\end{array}\right)$$

The functions $h_{kp}(t)$ are the normalized version of $\sin(\pi t)$ and $\cos(\pi t)$ for $p = 1$, $t^2 - 13t + 43/6$ and $t + 0.5$ for $p = 2$, and e^t, $t - (e-1)^{-1}$ and $t^2 - 1.0025t + 0.1654$ for $p = 3$. Then we transform $Q_{ip}(t)$ to

$$R_{ip}(t) = \frac{Q_{ip}(t)}{\sigma_p(t)}, \text{ where } \sigma_p^2(t) = \sum_{k=1}^{K_p} \text{var}(Z_{ipk}) h_{kp}^2(t) + \tau_p^2, \tag{1.22}$$

so that $R_{ip}(t)$ has a marginal N(0,1) distribution.

2. Generate the observations

$$Y_{ip}(t) = \mu_p(t) + \sigma_p(t) G^{-1}\left[\Phi\left\{R_{ip}(t)\right\}; \alpha_p(t)\right], \tag{1.23}$$

where $\Phi(\cdot)$ and $G(\cdot; \alpha)$ denote the cumulative distribution functions of the standard normal and skew normal distribution with mean 0, variance 1 and shape parameter α, respectively. The mean and shape parameter functions are

$$\begin{aligned} &\mu_1(t) = 5, && \alpha_1(t) = 0 \\ &\mu_2(t) = -3t^5 + 4.5t^3 - 1.5t + 15, && \alpha_2(t) = 10\sin(2\pi t) \\ &\mu_3(t) = 10e^{-t^2/2} + 20t^2 + 10, && \alpha_3(t) = 48t^2 - 48t + 6. \end{aligned} \tag{1.24}$$

The standard deviation function is given in equation (1.22).

The plots of simulation results are shown in Figures 1.2-1.5. Figure 1.2 shows the estimates of the mean (first column), standard deviation (second column) and shape parameter (third column) functions of each outcome from all 100 simulated datasets. The black lines are the true values, and the gray lines are the estimates. We assumed the skew normal distribution. These estimates were obtained by maximizing the pointwise likelihood at each point t_j using `R` package `sn` (Azzalini, 2011). The estimates were then smoothed further using penalized truncated cubic polynomial splines. The smoothing parameters were selected

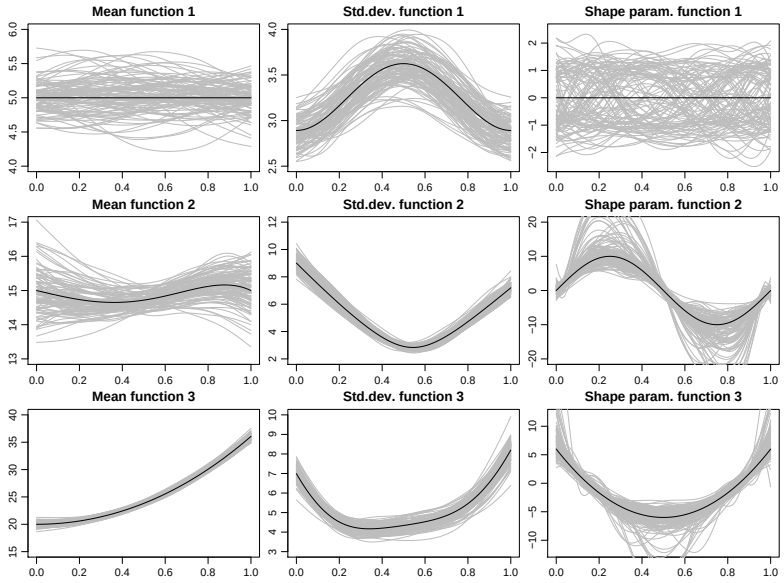

Figure 1.2: The estimates of mean, standard deviation and shape parameter functions for the simulated data.

using the REML criterion implemented in R package `mgcv` (Wood, 2006). Even though there seems to be high variation in the estimates when $\alpha \approx 10$, the density of skew normal for α equal to 10 is not much different from when α is much higher. This means that even though the estimate of α in this range is not very accurate, we can still obtain a good estimate for the distribution.

The dependence among the Gaussian copula were estimated using the orthonormal transform of cubic splines with 9 knots. The number of components and smoothing parameters were chosen based on AIC. The transformation and parameter selection are explained in Section 1.3.2. Figure 1.3 shows the true and estimates of the covariance from 2 datasets. The true covariances within an outcome and cross covariances between outcomes are shown in the first and fourth rows, respectively. The estimated covariances within an outcome from 2 datasets are shown in rows 2 and 3. The estimated cross covariance between outcomes from the same 2 datasets are shown in rows 5 and 6. Thus, rows 2 and

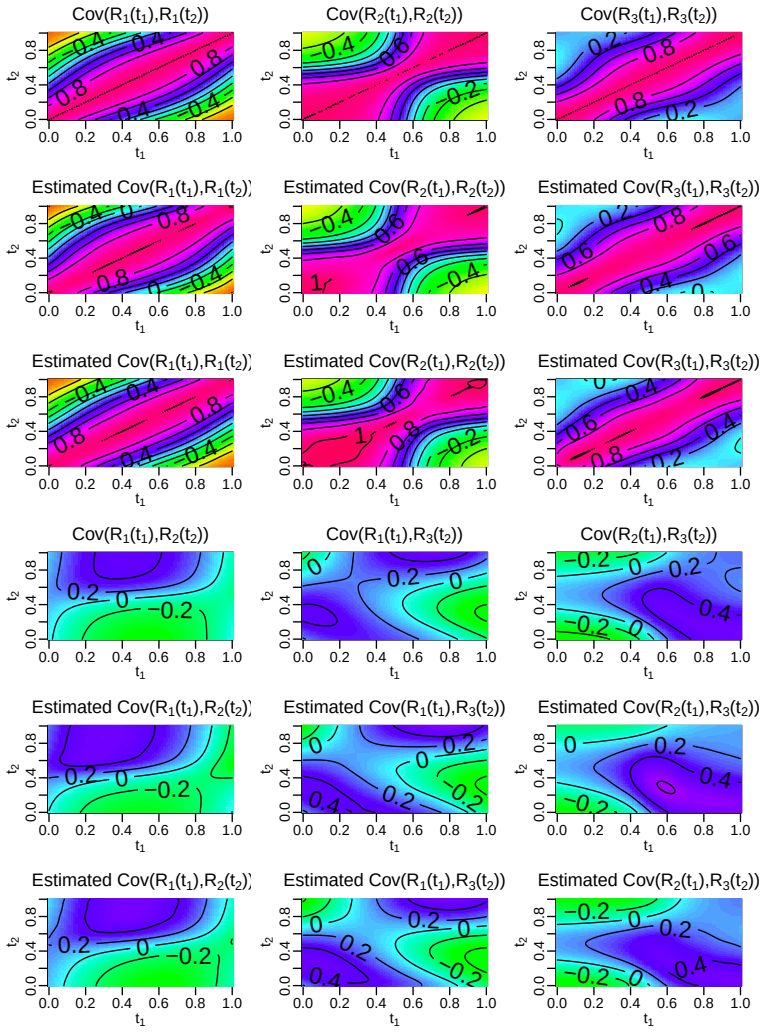

Figure 1.3: True and estimated covariances within each outcome and cross-covariances between outcomes.

3 should be compared with row 1 and rows 5 and 6 with row 4.

Estimated variances of each latent process are shown in Figure 1.4. The values close to 1, which are shown as black, red and blue lines, indicate the closeness to our assumption that the latent processes have marginal variance of 1. The square root of the integrated mean square error (IMSE), integrated square bias (IBIAS) and integrated variance (IVAR) for the marginal parameter functions and covariance parameters are shown in Table 1.1. These quantities are

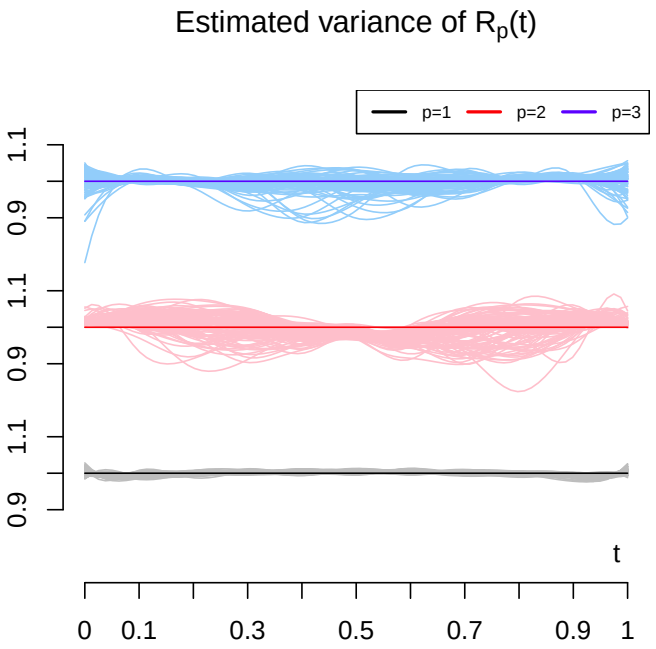

Figure 1.4: Estimates of the variance of the latent processes for the simulated data.

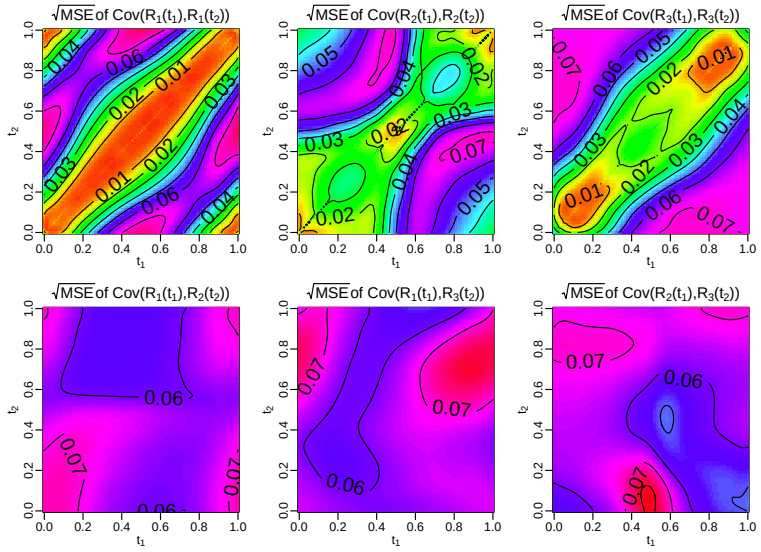

Figure 1.5: Pointwise square root of MSE for covariance within and between outcomes.

Table 1.1: Estimates of the square roots of IMSE, IBIAS and IVAR for the simulated data from 100 datasets

Parameter	$\sqrt{\text{IMSE}} \times 10^2$	$\sqrt{\text{IBIAS}} \times 10^2$	$\sqrt{\text{IVAR}} \times 10^2$
$\mu_1(t)$	23.66	2.51	23.52
$\mu_2(t)$	36.87	1.90	36.83
$\mu_3(t)$	36.66	3.94	35.84
$\sigma_1(t)$	16.72	2.20	16.58
$\sigma_2(t)$	29.24	1.78	29.19
$\sigma_2(t)$	28.57	3.36	28.37
$\alpha_1(t)$	98.00	6.93	97.76
$\alpha_2(t)$	210.7	72.42	197.9
$\alpha_2(t)$	105.6	16.71	104.2
$\text{cov}(R_1(t), R_1(t))$	4.25	0.43	4.23
$\text{cov}(R_2(t), R_2(t))$	4.58	0.57	4.55
$\text{cov}(R_3(t), R_3(t))$	4.52	0.73	4.46
$\text{cov}(R_1(t), R_2(t))$	6.30	0.64	6.27
$\text{cov}(R_1(t), R_3(t))$	6.49	0.98	6.42
$\text{cov}(R_2(t), R_3(t))$	6.36	0.66	6.32

calculated, for example, for the mean function by IMSE = $\int_0^1 \widehat{E}\{\widehat{\mu}(t) - \mu(t)\}^2 dt$, IVAR = $\int_0^1 \widehat{\text{Var}}\{\widehat{\mu}(t)\} dt$ and IBIAS = $\int_0^1 \left[\widehat{E}\{\widehat{\mu}(t)\} - \mu(t)\right]^2 dt$. Here $\widehat{E}$ and $\widehat{\text{Var}}$ denote the sample mean and sample variance, respectively. Overall, the bias and variance are low for the mean and standard deviation estimates. The bias for the shape parameter estimates are somewhat high, especially, as seen from Figure 1.2, at locations with high α (in absolute value). This should not affect our estimated marginal distributions too much as explained earlier. The covariance estimates have small IMSE. The contour plots of pointwise square root of the

mean square error for the covariance estimates are shown in Figure 1.5.

1.5 Discussion

We have developed a copula-based approach for modeling functional data with multiple outcomes. The copula approach, together with the KL expansion and low-rank spline model, significantly reduce the dimension of the model and allow fast estimation of the marginal distributions and the dependence structure. This is essential in the analysis of functional data, especially in the case of multiple outcomes where the data set can be very large.

Our methodology performed very well in our simulation study. Both of the marginal distributions and the dependence structure were estimated very well. In the following chapter, the methodology is applied to the three outcomes from the DTI study. This methodology can successfully identify the difference in both the marginal distributions and the dependence structure between the MS and healthy control groups.

One possible extension to our Gaussian copula assumption would be a t-copula. The t-copula covers a larger class of models than Gaussian copula, and is better in capturing tail dependence behavior (McNeil et al., 2005). The KL expansion used in our model does not apply directly to the t-copula. Another efficient alternative methodology will need to be developed.

CHAPTER 2

DIFFUSION TENSOR IMAGING STUDY APPLICATIONS

2.1 Introduction

In this section, we apply our methodology to the diffusion tensor imaging (DTI) data. DTI is a magnetic resonance imaging technique that measures the diffusion of water in tissue. The anisotropy in water diffusion allows the images of the white matter in the brain to be generated. White matter tracts are made up of axons that transmit signals between different regions of the brain. These axons are surrounded by fatty substance called myelin that helps the signal to be carried quickly. Multiple sclerosis (MS) is an autoimmune disease that is caused by the damage to myelin and can lead to significant disabilities in patients (Goldsmith et al., 2012).

There have been many other studies on the DTI tractography for MS patients. For instance, Reich et al. (2005) analyzed various indices obtained from DTI tractography to identify abnormalities in MS patients. Goldsmith et al. (2011a,b, 2012) studied DTI data using the penalized functional regression approach. Longitudinal functional PCA (LFPCA) was introduced by Greven et al. (2010) to account for the longitudinal variability of tractography data in MS. LFPCA extended the MFPCA approach of Di et al. (2009) who focused on replicated functional data. See Staicu et al. (2010) for the application of PCA for complex multilevel spatially correlated functional data. McLean et al. (2012) introduced a functional generalized additive model and applied it in the study of the relationship between a cognitive test score and DTI tractography.

DTI provides many measurement of water diffusion. In this study, we consider 3 measurements: 1) fractional anisotropy, 2) parallel diffusivity and 3) perpendicular diffusivity. Parallel and perpendicular diffusivities are diffusion parallel and perpendicular to the long axis of a fiber bundle, respectively, while fractional anisotropy measures the difference of diffusion in the two directions. Mathematically, at each location, DTI tractograpy can be described as a 3×3 symmetric, positive definite matrix. Suppose the eigenvalues of the matrix are $\lambda_1 > \lambda_2 > \lambda_3 > 0$. The parallel and perpendicular diffusivities are given by λ_1 and $\frac{1}{2}(\lambda_2 + \lambda_3)$, respectively. The fractional anisotropy is calculated as

$$\left[\frac{3\left\{(\lambda_1 - \bar{\lambda})^2 + (\lambda_2 - \bar{\lambda})^2 + (\lambda_3 - \bar{\lambda})^2\right\}}{2(\lambda_1^2 + \lambda_2^2 + \lambda_3^2)}\right]^{1/2}, \tag{2.1}$$

where $\bar{\lambda} = (\lambda_1 + \lambda_2 + \lambda_3)/3$ (McLean et al., 2012).

We demonstrate our methodology using the DTI data from the healthy and MS groups. The objective is to identify the locations where the abnormalities occur and also explain the characteristics of the abnormalities. The intention is to use differences between controls and diseased subjects in any DTI characteristics to help understand the disease process. These differences should also be useful in detecting MS and also in monitoring the progress of the the treatment.

Parameter estimation for the three outcomes for DTI data is explained in Section 2.2. Staicu et al. (2011) studied parallel diffusivity for the same dataset and found that all of the mean, variance and skewness are different between the two groups. We found the same results for parallel diffusivity. We did an additional study on fractional anisotropy and perpendicular diffusivity and found that only the mean and variance are significantly different. The correlations within each outcome are similar for the two groups. Our most important finding is that the cross-correlation between fractional anisotropy and parallel

diffusivity is different between the two groups in the middle part of the tract (roughly from tract locations 0.2 to 0.8). The cross correlation in this tract section is slightly positive in the control healthy group, while it is negative in the MS group. Our bootstrapping result confirms that the difference is significant. The difference between the two groups of the cross-correlation between the fractional anisotropy and the perpendicular diffusivity is less apparent, while the cross-correlation between the perpendicular and parallel diffusivities is similar in most parts of the tract.

In Section 2.3, we introduce an approach for MS case prediction utilizing the difference we found in the marginal distribution and the correlation between the MS and healthy control groups. The receiver operating characteristic (ROC) curves show that our approach can potentially be used to predict the MS case status for patients. This chapter concludes with discussion in Section 2.4

2.2 Parameter Estimation for DTI Data

In this section, we apply our methodology to the DTI data. A subset of our data set is freely available as the `DTI` data set in `R`'s `refund` package Crainiceanu et al. (2012).

Our data consists of measurements along the corpus callosum from 42 healthy controls and 162 MS patients, measured at 93 locations along the tract. Figure 1.1 in Section 1.1 displays the 3 outcomes for both groups. Our goal is to estimate, for each group, the marginal distributions of each outcome, the dependence structure across locations within each outcome and across different outcomes, and to identify the difference between the two groups.

2.2.1 Marginal Parameter Estimation for DTI Data

We assumed that each measurement from each group follows a skew-t marginal distribution. We assumed that the mean, standard deviation and shape parameter vary along the tract, while the degrees of freedom are constant along the tract locations. Using `R` package `sn` (Azzalini, 2011), we obtained the maximum likelihood estimates for the degree of freedom for the fractional anisotropy, parallel diffusivity and perpendicular diffusivity equal to 24.93, 11112 and16.30, respectively for the control group, and 11732, 5.98 and 8.47 for the MS group. We used the skew normal distribution to model any outcome with the estimated degrees of freedom > 100. The undersmoothed estimates for the mean, standard deviation and shape parameters were obtained via pointwise maximum likelihood estimation using function `st.mle` in the `R` package `sn` (Azzalini, 2011). The mean and standard deviation estimates were smoothed further by penalized spline fitting using generalized additive model implemented in package `mgcv` in `R` (Wood, 2006). The shape parameters were smoothed further by fitting the penalized spline to maximize the penalized pseudo likelihood computed using the estimates of the mean and standard deviation as in equation (1.4).

The estimated mean, variance and skewness for the three outcomes for both groups are shown in Figure 2.1. The estimates for the control group are shown in red lines and the estimates for the MS group are shown in black lines. The gray dashed lines indicate zero skewness. Figure 2.2 shows the estimated differences of the parameter functions between the two groups, which are shown in black solid lines. The black dashed lines show 90% pointwise confidence intervals computed using bootstrapping with 1000 samples. The gray dashed

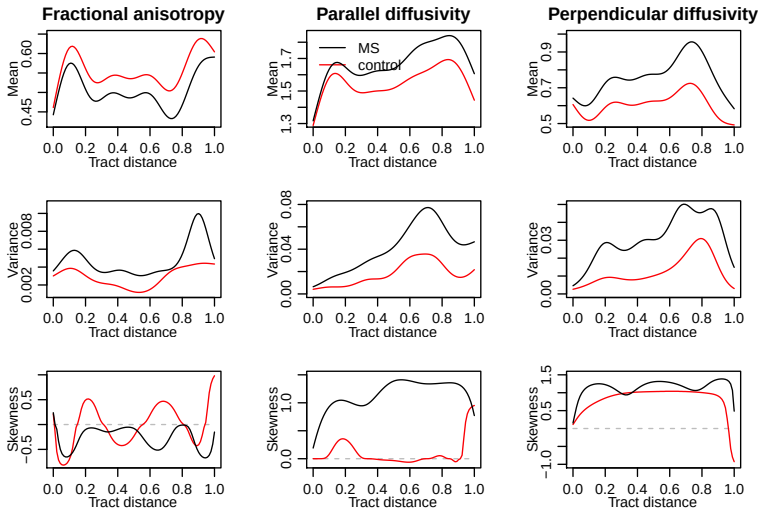

Figure 2.1: Estimates of the mean, variance and skewness for the healthy control and MS groups.

horizontal lines are through 0. The mean, variance and skewness functions are all statistically significantly different between the two groups for parallel diffusivity, while only the mean and variance are significantly different for fractional anisotropy and perpendicular diffusivity. The mean of fractional anisotropy for the MS group is lower than the control group, but the means of parallel and perpendicular diffusivities are higher in the MS group. The variances for the MS group are higher in all of the three outcomes. The skewness of the fractional anisotropy is close to zero for both groups. The parallel diffusivity of the MS group is positively-skewed while the skewness is close to zero in most part of the tract in the control group. The difference in the skewness between the two groups is significant. The perpendicular diffusivity is positively-skewed in most part of the tract in both groups. There is no significant difference in the skewness of perpendicular diffusivity between the two groups.

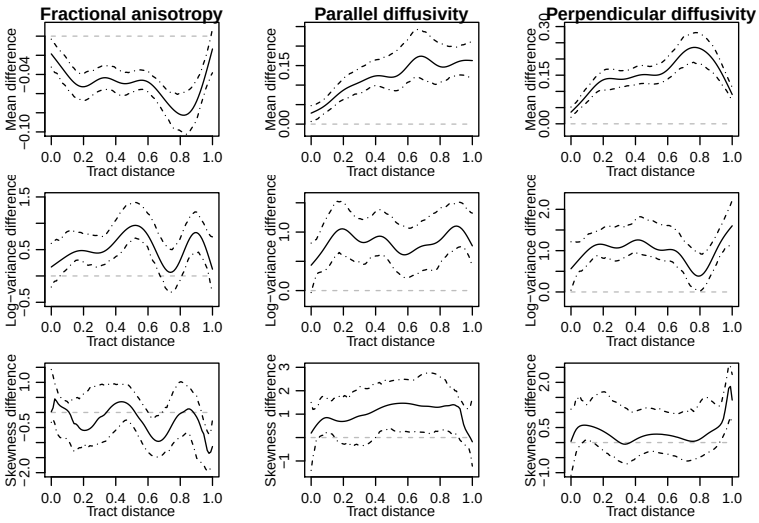

Figure 2.2: Estimated differences in the mean, log-variance and skewness between the healthy and MS groups.

2.2.2 Dependence Structure of DTI Data

After the estimates $\widehat{\mu}_p, \widehat{\sigma}_p, \widehat{\alpha}_p$ for the mean, standard deviation and shape parameters for the outcome $p = 1, 2, 3$ for each group were obtained, the data were transformed using equation (1.5). Here, G is the skew-t distribution with mean 0, variance 1, shape parameter $\widehat{\alpha}_p$ and degree of freedom for each outcome and each group as stated earlier. The method explained in Section 1.3 was applied to the transformed data. The transformation implies that the marginal variance of the transformed process is 1, and, hence, the covariance function for the transformed process is also its correlation function. In our study, estimates of the marginal variance vary from 0.6 to 1.3 for the MS group, and 0.6 to 1.5 for the control group, due to the small sample sizes. We verified this by a simple simulation as follows. We generated data with 42 observations for control group, and 162 observations for MS group according to the skew t distribution using the estimated parameters obtained earlier. Then we transformed the simulated data as in Equation (1.5) and computed the marginal sample variances.

The variances vary similarly to those of the DTI data. The sample variances are noticeably closer to 1 as the number of observations for each group increases to 200. Most of the sample variances lie well within [0.9,1.1] as the number of observations increases to 500.

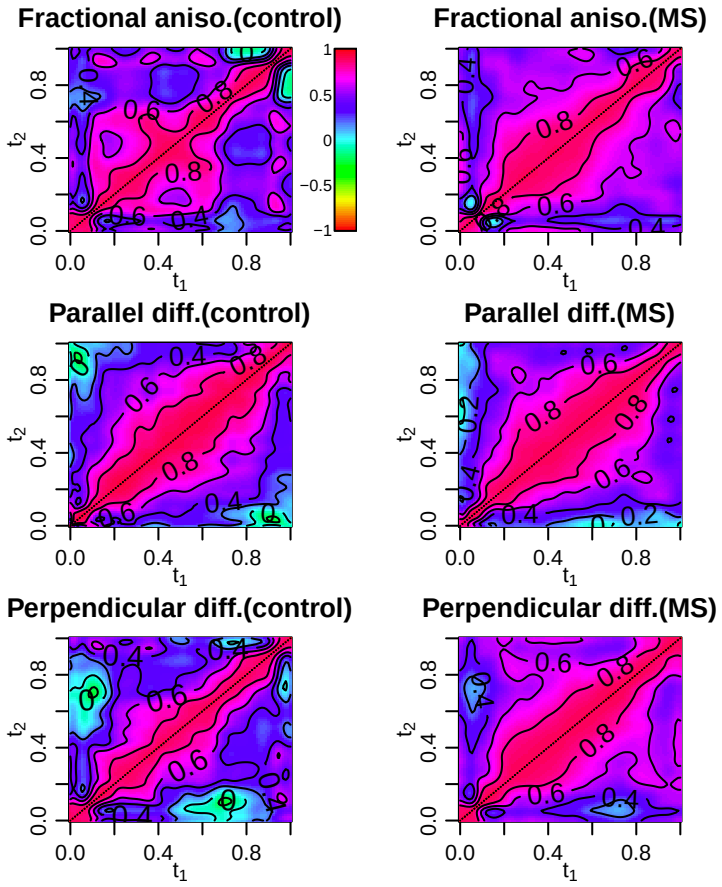

Figure 2.3: Estimated correlations within each outcome for the control and the MS groups.

The estimated correlations for the DTI data were obtained by converting the estimated covariances for the transformed process (i.e., scale the covariances by the marginal standard deviations). Figure 2.3 displays the estimated correlations within each of the three outcomes for the healthy and MS groups. Fig-

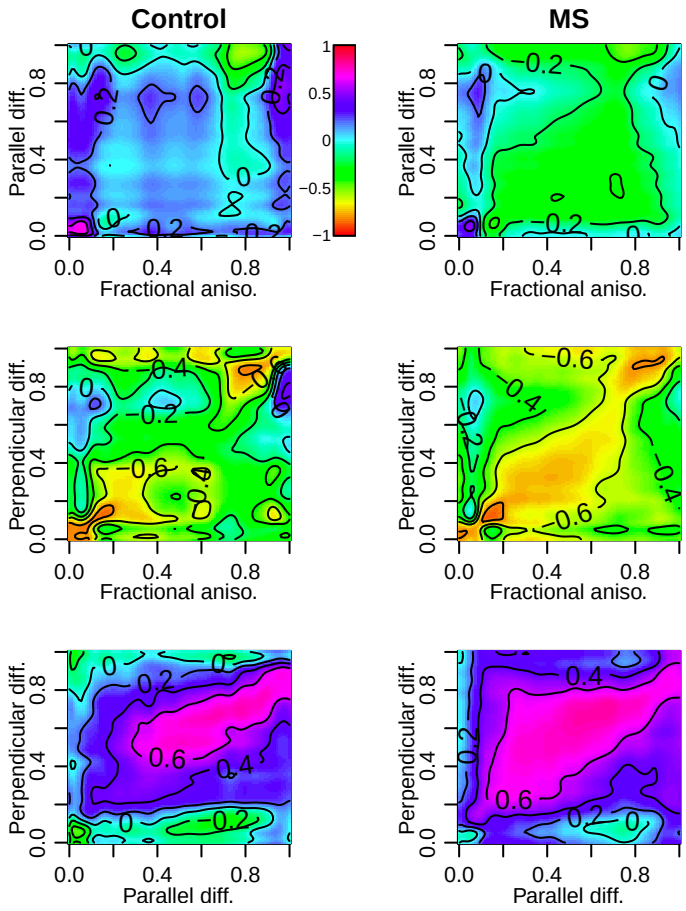

Figure 2.4: Estimated cross-correlations between different outcomes for the control and MS groups.

ure 2.4 displays the estimated cross-correlations between different outcomes for the control and MS groups. Both figures show the estimates for the control group on the left panel and for the MS group in the right panel. The correlations within the outcomes look somewhat similar for the two groups, while there is a noticeable difference between the two groups for the cross-correlations, especially for the cross-correlation between parallel diffusivity and fractional anisotropy. The pointwise confidence intervals for the estimated correlations can be computed using bootstrapping. Figure 2.5 shows the significance levels

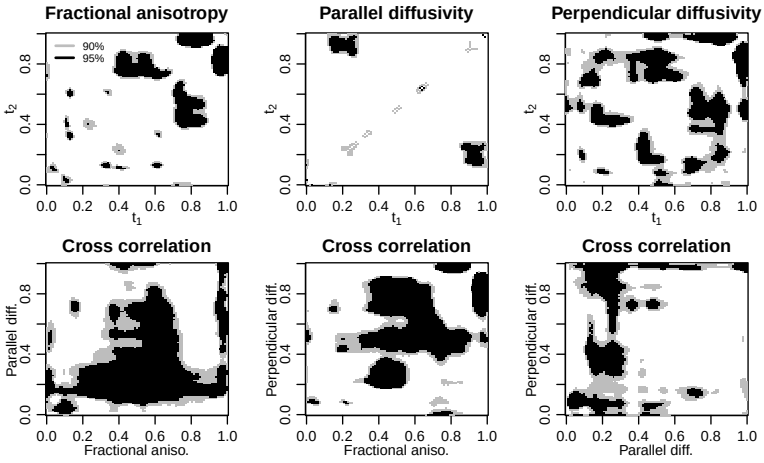

Figure 2.5: The differences of the correlations within each outcome (top row) and between outcomes (bottom row).

of the difference of the correlations between the two groups calculated from the 90% and 95% bootstrap pointwise confidence intervals based on 1000 samples. The top row shows the plots for correlation within the same outcome, while the bottom row shows the those across different outcomes. The gray and black regions are the areas at which the differences are statistically significant at 90% and 95% levels, respectively. The computation is based on bootstrap pointwise confidence intervals using 1000 samples.

The bootstrapping results confirm that there is no significant difference in the correlations within fractional anisotropy and parallel diffusivity. The correlation within perpendicular diffusivity is different in various small regions of the tract. There is a significant difference between the two groups in the cross-correlation between the fractional anisotropy and the parallel diffusivity in a large middle part of the tract (at around locations 0.2 to 0.8). The cross-correlation is slightly positive in the control group, while it is negative in the MS group. The difference in the cross-correlation between the fractional anisotropy

and the perpendicular diffusivity is significant in a smaller part of the tract (at around locations 0.4 to 0.8). The cross-correlation is more negative in the MS group in this region. The cross-correlation between the parallel and perpendicular diffusivities is somewhat similar between the two groups with the parallel diffusivity at around location 0.2 being more positively-correlated with the overall perpendicular diffusivity in the MS group.

2.3 Case Status Prediction for DTI Data

In this section, we demonstrate an application of the prediction introduced in Section 1.3.4 to predict subjects' case status. We introduce a measure that utilizes the difference in both of the marginal distributions and the correlation structure between the two groups we found in the previous section. For each subject with unknown case status, we use the parallel and perpendicular diffusivities to obtain two predictions for the fractional anisotropy using 2 sets of parameters: one for the MS group and the other for the control group. If this subject is a MS patient, the prediction using the MS parameters should be more similar to the real observed curve than the prediction using the control parameters. One possible criterion for measuring the similarity between curves is to compare the areas between each predicted curve and the observed curve. The prediction can be similarly applied to the parallel and perpendicular diffusivities.

We studied the discriminating power of our prediction using a leave-one-out analysis. To predict case status for subject i, we obtained estimates for marginal and dependence parameters using all of the subjects but subject i. Then we performed prediction of an outcome using the other two outcomes as explained

in Section 1.3.4 to subject i using the obtained parameter estimates for MS and control groups. Let $\text{Area}_{i,\text{control},p}$ and $\text{Area}_{i,\text{MS},p}$ denote the areas between the observed outcome p of subject i and the predicted outcome p for subject i using the control and MS parameters, respectively. To predict the case status, we use a test statistic

$$T_{i,p} = \text{Area}_{i,\text{control},p} - \text{Area}_{i,\text{MS},p},$$

and classify this subject to be MS patient if $T_{i,p} \geq \tau_p$ for some threshold value τ_p. Figure 2.6 depicts receiver operating characteristic curves as the threshold values vary using predicted curves for each of the 3 outcomes: fractional anisotropy, parallel diffusivity and perpendicular diffusivity. The vertical axis shows the true positive rate which is the proportion of MS patients that are correctly classified, and the horizontal axis shows the false positive rate which is the proportion of healthy subjects that are incorrectly classified. The ROC curves and their confidence intervals were produced by the `R` package `pROC` (Robin et al., 2011). The 95% confidence intervals were computed using 10000 stratified bootstrap replicates. The gray horizontal lines show the confidence intervals for the false positive rates conditioning on the true positive rates, and vice versa for the vertical lines. The confidence intervals lie almost entirely above the 45° line, which indicates that the differences in correlation structure between the control and MS groups have statistically significant predictive power.

2.4 Discussion

In this chapter, we applied the copula-based approach introduced in the previous chapter to the three outcomes from the DTI study. We were able to identify the differences in the pointwise marginal distributions between the control and

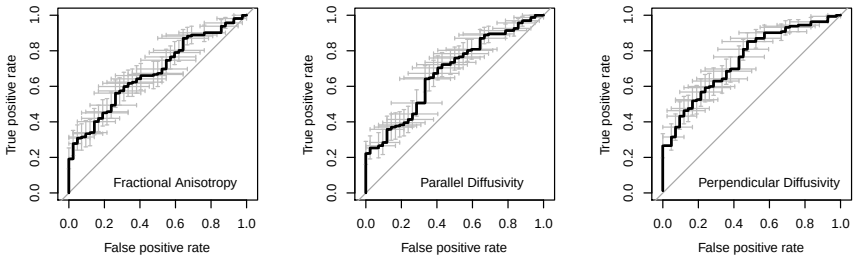

Figure 2.6: Receiver operating characteristic curves for case status prediction using the predicted fractional anisotropy, parallel diffusivity and perpendicular diffusivity.

MS groups. The mean of fractional anisotropy for the MS group is lower than the control group, while the means of parallel and perpendicular diffusivities for the MS group are higher than the control group. This is as expected due to the loss of myelin surrounding the white matter tracts in MS patients. The marginal variances for all of the three outcomes for the MS group are higher than those of the control group. The skewness seems to be statistically significantly different between the two groups only for the parallel diffusivity.

Our dependence structure study shows that the correlations within the same outcome are not different between the two groups, while the cross correlation between outcomes, especially between the parallel diffusivity and fractional anisotropy, are different. We used the difference we found in the marginal distributions and cross-correlations between outcomes to derive a measure to predict MS case status for patients. ROC curves show that the measure we derived is predictive of MS status.

CHAPTER 3

DIFFUSION TENSOR IMAGING STUDY IN LONGITUDINAL SETTINGS

3.1 Introduction

In the past chapters, the focus of the study is on identifying the difference of the diffusion tensor imaging (DTI) scan between multiple sclerosis (MS) patients and healthy controls. This includes both the pointwise marginal distribution of the DTI outcomes and the correlation across different outcomes. The three outcomes we have considered are fractional anisotropy, parallel diffusivity and perpendicular diffusivity which measure water diffusion in different directions as explained in Section 2.1. In this chapter, our focus is shifted toward understanding how the DTI outcomes of MS patients change over time. We consider the DTI data from 162 MS subjects. These are the same MS subjects we consider in Chapter 2. While the data we considered earlier contains only data from the first visit, the data considered in this chapter includes the outcomes from multiple visits of these patients. The number of visits for each subject varying from 1 to 8. The total number of observations is 421.

There have been studies on longitudinal DTI data. Some examples are Goldsmith et al. (2012) and Greven et al. (2010). Goldsmith et al. (2012) studied the relation between cognitive disability in MS patients by incorporating DTI tractography measurement from multiple visits as functional covariates in a longitudinal penalized functional regression setting. Greven et al. (2010) developed a longitudinal functional principal component analysis approach to decompose the longitudinal DTI data into a time-dependent population average, baseline subject-specific variability, longitudinal subject-specific variability,

subject/visit-specific variability and measurement error.

Let $Y_i(t, T_{ij})$ denote the data from an outcome, which can be fractional anisotropy, parallel diffusivity or perpendicular diffusivity of subject i observed at time T_{ij} at tract location t. Since the distribution of the data should depend on the time between visits instead of the order of the visits, instead of using j as the index for the visits of each subject, we use j as the count of 3-month intervals since the subject's first visit. In our DTI data, j can range from 0 to 17. As a result, we can shorten the notation $Y_i(t, T_{ij})$ to $Y_{ij}(t)$. Here j is rounded to the closest integer.

Similar to Chapters 1 and 2, we use a copula approach to model the marginal distributions and the dependence structure separately. The difference in the first step here is that we allow the marginal mean functions to vary between visits. That is, $Y_i(t, T_{ij})$ can be written as

$$Y_{ij}(t) = \mu_j(t) + \sigma(t)G^{-1}\left\{U_{ij}(t); \alpha(t)\right\} \tag{3.1}$$

Here, $\mu_j(t)$ are the marginal mean functions which can vary across j. $\sigma(t)$ and $\alpha(t)$ are the marginal standard deviation and shape parameter functions, respectively, which are assumed to be common across j. Here G can be, for example, a skew-normal or a skew-t distribution, with mean 0 and standard deviation 1. This model seems to assume that each subject starts visits at the same point in the disease process which is not the case in our DTI data. However, when we model the copula, the transformed process, which is a function of $U_{ij}(t)$, will be decomposed. The first random component is subject-specific, which should capture the variation across subjects prior to their first visit.

We proceed the marginal parameter estimation step similarly to the previous chapters. First, the maximum likelihood estimation is used at each location

t to obtain undersmoothed estimates for $\mu_j(t)$, $\sigma(t)$ and $\alpha(t)$. Then, these undersmoothed estimates are smoothed further using one-dimensional penalized splines for $\sigma(t)$ and $\alpha(t)$, and two-dimensional penalized splines for $\mu_j(t)$. After the estimates $\widehat{\mu_j}(t)$, $\widehat{\sigma}(t)$ and $\widehat{\alpha}(t)$ for $\mu_j(t)$, $\sigma(t)$ and $\alpha(t)$ are computed, the observed data $Y_{ij}(t)$ can be transformed as

$$R_{ij}(t) = F_t^{-1}\left[G\left\{\frac{Y_{ij}(t) - \widehat{\mu_j}(t)}{\widehat{\sigma}(t)}; \widehat{\alpha}(t)\right\}\right] \tag{3.2}$$

where F_t is the CDF of $R_{ij}(t)$. In the DTI study, we assume a Gaussian copula; that is F_t^{-1} is the CDF of standard normal distribution. We apply a model for modeling spatially correlated multilevel functional data proposed by Staicu et al. (2010) to model the transformed processes $R_{ij}(t)$. The key idea to estimate the correlation of $R_{ij}(t)$ is to decompose the data into 1) subject-specific part, 2) second-level visit-specific deviation from the subject-specific mean, and 3) noise.

In Section 3.2, we explain the estimation for the marginal parameters and the estimation for each component of $R_{ij}(t)$. In Section 3.3, we present results from DTI data. This chapter concludes with discussion in Section 3.4

3.2 Estimation Methods

Let $Y_{ij}(t, T_{ij})$ denote the data of an outcome of subject i, $i = 1, 2, \ldots, N$ observed at location t and at time T_{ij}. To simplify the notation, assume that all subjects have common observed time $T_1, \ldots, T_J$; hence we can shorten the notation $Y_{ij}(t, T_{ij})$ to $Y_{ij}(t)$. Our DTI data doesn't have common observed time; however, the computation for the case of different observed time can be easily generalized. Using the copula approach, as explained in Section 1.2, we can write $Y_{ij}(t)$ as in

equation (3.1). In Section 3.2.1, we explain the estimation of the marginal parameter functions. The standard deviation and shape parameter functions can be generalized to vary across j when appropriate, as long as the number of observations is adequate. The methodology we present for $\mu_j(t)$ inSection 3.2.1 can be easily applied to $\sigma(t)$ and $\alpha(t)$. Caution should be taken when the skew t distribution is used as the shape parameter depends on the degree of freedom for which the estimation is highly sensitive especially in the case of small number of observations. Once the estimates $\left\{\widehat{\mu}_j(t) : j = 1, 2, \ldots, J\right\}$, $\widehat{\sigma}(t)$ and $\widehat{\alpha}(t)$ for $\left\{\mu_j(t) : j = 1, 2, \ldots, J\right\}$, $\sigma(t)$ and $\alpha(t)$ are obtained, the data $Y_{ij}(t)$ are transformed using equation (3.1). The first part of this section explains the estimation for the marginal parameters. The second part presents the estimation for each component of $R_{ij}(t)$.

3.2.1 Marginal Parameter Estimation

The estimation for the marginal parameters is done in 2 steps, similar to Section 1.2.1. In the first step, undersmoothed estimates for these parameters are obtained; then the estimates are smoothed further in the second step to obtain the final estimates.

For simplicity, suppose the data are observed on a common dense grid of points $t_{ik} = t_k$ for all i, j and $k = 1, 2, \ldots, m$. We first obtain undersmoothed estimates $\widetilde{\sigma}(t)$ and $\widetilde{\alpha}(t)$ for $\sigma(t)$ and $\alpha(t)$ by maximizing the pointwise likelihood function for each point t_k

$$\sum_{i=1}^{N} \log\left[g\left\{\frac{Y_i(t_k) - \mu_j(t_k)}{\sigma(t_k)}\right\}; \alpha(t_k)\right] - \log\{\sigma(t_k)\} \tag{3.3}$$

where $g(y; \alpha) = \partial G(y; \alpha)/\partial y$ is the density function corresponding to the distribu-

tion function G. When G is assumed to be a skew normal or skew-t distribution, the estimates $\widetilde{\mu}_j(t)$, $\widetilde{\sigma}(t)$ and $\widetilde{\alpha}(t)$ can be computed using the functions `sn.mle` or `st.mle` from `R` package `sn` (Azzalini, 2011). Note that usually in this step, in addition to $\widetilde{\sigma}(t)$ and $\widetilde{\alpha}(t)$, we also obtain estimates for $\widetilde{\mu}_j(t)$ which we will ignore and consider only $\widetilde{\sigma}(t)$ and $\widetilde{\alpha}(t)$. After $\widetilde{\sigma}(t)$ and $\widetilde{\alpha}(t)$ are obtained, we calculate the pointwise maximum likelihood estimates $\widetilde{\mu}_j(t)$ for each j by maximizing

$$\sum_{i=1}^{N} \log\left[g\left\{ \frac{Y_i(t_k) - \mu_j(t_k)}{\widetilde{\sigma}(t_k)} \right\}; \widetilde{\alpha}(t_k) \right] - \log\{\widetilde{\sigma}(t_k)\}. \tag{3.4}$$

When a skew t or skew normal distribution is assumed, the functions `dst` and `dsn` from `R` package `sn` (Azzalini, 2011) can be used to compute the density function g, aiding in the maximization step.

Now the undersmoothed estimates $\left\{\widetilde{\mu}_j(t) : j = 1, 2, \ldots, J\right\}$, $\widetilde{\sigma}(t)$ and $\widetilde{\alpha}(t)$ can be smoothed further. This step is done for $\widetilde{\sigma}(t)$ and $\widetilde{\alpha}(t)$ using one-dimensional penalized spline as in step 2 of Section 1.2.1 to obtain the final estimates $\widehat{\sigma}(t)$ and $\widehat{\alpha}(t)$

The smoothing of $\left\{\widetilde{\mu}_j(t)\right\}_{j=1,2,\ldots,J}$ is done using two-dimensional thin-plate splines. Let stack $\widetilde{\mu}_j(t)$ across j and call the resulting function of t and j $\widetilde{\mu}$. The goal is to find $\widehat{\mu}$ that minimizes

$$||\hat{\mu} - \tilde{\mu}||^2 + J_m(\hat{\mu}) \tag{3.5}$$

where

$$J_m(\hat{\mu}) = \int\int \sum_{\alpha_1+\alpha_2=m} \frac{m!}{\alpha_1!\alpha_2!} \left(\frac{\partial^m \widehat{\mu}}{\partial t^{\alpha_1} \partial j^{\alpha_2}} \right)^2 dtdj \tag{3.6}$$

with $m > 1$. See Duchon (1975, 1976, 1977) and Meinguet (1979) for theoretical foundations for the thin-plate spline. Wahba (1990) and Green and Silverman (1994) provide further explanation about thin-plate splines. Additional details about low-rank thin-plate regression splines (TPRS), which are low-rank

approximations of thin-plate splines, can be found in Wood (2003). Some of `R` functions that provide thin-plate spline fitting are function `Tps` in package `fields` (Furrer, Nychka, and Sain, 2013) and function `gam` (which computes a TPRS) in package `mgcv` (Wood, 2006). Tensor product splines can also be used to fit the marginal mean functions and can also be computed using function `gam`. After the smoothing step, we now obtain the final estimates for the marginal parameters.

3.2.2 Estimation of the Dependence Structure

In this study, we assume a Gaussian copula. The data is transform using the estimates from the previous section as

$$R_{ij}(t) = \Phi^{-1}\left[G\left\{\frac{Y_{ij}(t) - \widehat{\mu}_j(t)}{\widehat{\sigma}(t)}; \widehat{\alpha}(t)\right\}\right] \tag{3.7}$$

where Φ is the cumulative distribution function of the standard normal distribution, and G is as in the previous section. As explained in Section 3.1, we can decompose $R_{ij}(t)$ as

$$R_{ij}(t) = V_i(t) + Q_{ij}(t) + \epsilon_{ij}(t), \tag{3.8}$$

where $V_i(t)$ is the subject-specific part, $Q_{ij}(t)$ is the second-level visit-specific deviation from the subject-specific mean and $\epsilon_{ij}(t)$ are white noise processes. We assume that each process is Gaussian and has zero mean. Note that $Q_{ij}(t)$ are correlated across j and t. Modeling $Q_{ij}(t)$ directly would create a large-dimensional problem. To reduce the model complexity, we further decompose $Q_{ij}(t)$ into 2 parts

1. the part that depends only on the tract location t, and exhibits no between-visit correlation $W_{ij}(t)$, and

2. the part that depends only on the visit time j and exhibits between-visit correlation S_{ij}.

This is a standard technique in multilevel model. We can rewrite $R_{ij}(t)$ as

$$R_{ij}(t) = V_i(t) + W_{ij}(t) + S_{ij} + \epsilon_{ij}(t). \tag{3.9}$$

$W_{ij}(t)$ and S_{ij} are assumed to be Gaussian. In this section, we apply the method for modeling spatially correlated multilevel functional data introduced by Staicu et al. (2010) to estimate each component of $R_{ij}(t)$. We make the following assumptions:

1. V_i and W_{ij} are square integrable, independent over i and j with mean 0 and some specific covariance functions.
2. $S_i = \left\{S_{ij} : j = 1, 2, \ldots\right\}$ is a second order stationary process with autocovariance function decreasing to zero. Let $\nu^s(\Delta)$ and $\rho^s(\Delta)$ denote the autocovariance and autocorrelation functions of S_i. Let Δ^* be a threshold such that $\rho^S(\Delta)$ is negligible for $\Delta \geq \Delta^*$.
3. V_i W_{ij} and ϵ_{ij} are mutually independent.

The processes V_i and W_{ij} have a unique representation, provided that there are no ties among their eigenfunctions,

$$V_i(t) = \sum_{k_1=1}^{\infty} \xi_{i,k_1}\phi_{k_1}^{(1)}(t) \text{ and } W_{ij}(t) = \sum_{k_2=1}^{\infty} \zeta_{ij,k_2}\phi_{k_2}^{(2)}(t) \tag{3.10}$$

where the random coefficients ξ_{ik_1} and ζ_{ijk_2} are given by

$$\xi_{ik_1} = \int V_i(t)\phi_{k_1}^{(1)}(t)dt, \text{ uncorrelated across } k_1, \tag{3.11}$$

$$\zeta_{ijk_2} = \int W_{ij}(t)\phi_{k_2}^{(2)}(t)dt, \text{ uncorrelated across } k_2, \tag{3.12}$$

and $\phi_{k_1}^{(1)}(t)$ and $\phi_{k_2}^{(2)}(t)$ are eigenfunctions of $K^V(t,s) = \text{Cov}(V_i(t), V_i(s))$ and $K^W(t,s) = \text{Cov}(W_{ij}(t), W_{ij}(s))$, respectively (Levy, 2008). We also have that $\text{var}(\xi_{ik_1}) = \lambda_{k_1}^{(1)}$, the eigenvalues of K^V, and $\text{var}(\zeta_{ijk_2}) = \lambda_{k_2}^{(2)}$, the eigenvalues of K^W. Since V_i and W_{ij} are independent, $\{\xi_{ik_1} : k_1 = 1, 2, \ldots\}$ are uncorrelated with $\left\{\zeta_{ijk_2} : k_2 = 1, 2, \ldots\right\}$. We estimate V_i and W_{ij} using finite truncation

$$V_i(t) = \sum_{k_1=1}^{K_1} \xi_{i,k_1}\phi_{k_1}^{(1)}(t) \text{ and } W_{ij}(t) = \sum_{k_2=1}^{K_2} \zeta_{ij,k_2}\phi_{k_2}^{(2)}(t), \tag{3.13}$$

where the numbers of components K_1 and K_2 need to be determined.

The following are the steps for the calculation for each component of $R_{ij}(t)$

1. Estimate autocovariance function of S_i, $\nu^S(\Delta)$.
2. Estimate covariance operators $K^V(t,s)$ and $K^W(t,s)$.
3. Obtain the estimated eigenvalues and eigenfunctions of K^V and K^W.
4. Estimate the principal component scores $\{\xi_{ik_1} : k_1 = 1, 2, \ldots\}$ and $\left\{\zeta_{ijk_2} : k_2 = 1, 2, \ldots\right\}$.
5. Estimate σ_ϵ^2

First, we estimate $\nu^S(\Delta)$ for $\Delta \geq 0$. Consider visit times j_1 and j_2 that are Δ apart; that is $\Delta = |j_1 - j_2|$. The between-visit covariance is

$$\begin{aligned} K_B^R(t,s,\Delta) &= \text{Cov}\left\{R_{ij_1}(t), R_{ij_2}(s)\right\} &(3.14)\\ &= K^V(t,s) + \nu^S(\Delta) &(3.15)\end{aligned}$$

Define

$$K_W^R(t,s,\Delta) = \frac{1}{2}\text{Cov}\left\{R_{ij_1}(t) - R_{ij_2}(t), R_{ij_1}(s) - R_{ij_2}(s)\right\} \tag{3.16}$$

which can be shown to equal to

$$K^W(t,s) + \sigma_S^2 - \nu^S(\Delta) + \sigma_\epsilon^2\delta_{ts} \tag{3.17}$$

where $\sigma_S^2 = v^S(0)$, σ_ϵ^2 is the variance of white noise $\epsilon_{ij}(t)$, and δ_{ts} is 1 if $t = s$ and 0 otherwise. $K_W^R(t, s, \Delta)$ will be estimated using method of moments, and $v^S(\Delta)$ can be estimated using

$$v^S(\Delta) \approx v^S(\Delta) - v^S(\Delta^*) = K_W^R(t, s, \Delta^*) - K_W^R(t, s, \Delta). \tag{3.18}$$

The value of Δ^* can be specified by experts in the field. In our DTI study, we restrict the range of possible Δ^* to be relative large to be safe and use crossvalidation to choose Δ^*.

First, consider the case $\Delta < \Delta^*$. $K_W^R(t, s, \Delta)$ is estimated by

$$\hat{K}_W^R(t, s, \Delta) = \frac{1}{2} \frac{\sum_{i,j} \sum_{j \neq j'} w_{jj'}(\Delta) \left\{R_{ij}(t) - R_{ij'}(t)\right\} \left\{R_{ij}(s) - R_{ij'}(s)\right\}}{\sum_{i,j} \sum_{j' \neq j} w_{jj'}(\Delta)} \tag{3.19}$$

where $w_{jj'}(\Delta)$ equal 1 if $|j - j'| \in \{\Delta - w, \ldots, \Delta - 1, \Delta, \Delta + 1, \ldots, \Delta + w\}$, and 0 otherwise. The width w can be selected using crossvalidation.

For the case $\Delta \geq \Delta^*$, $K_W^R(t, s, \Delta)$ is estimated by

$$\hat{K}_W^R(t, s, \Delta) = \frac{1}{2|N(\Delta^*)|} \sum_{i,j} \sum_{j' : |j' - j| \geq \Delta^*} \left\{R_{ij}(t) - R_{ij'}(t)\right\} \left\{R_{ij}(s) - R_{ij'}(s)\right\} \tag{3.20}$$

where $N(\Delta^*) = \{(i, j, j') : |j - j'| \geq \Delta^*\}$. Note that the right hand size doesn't depend on Δ. Now we use equation (3.18) and average over (t, s) to have

$$\hat{v}^S(\Delta) = \frac{1}{|(t, s) : t \leq s|} \sum_t \sum_{t \leq s} \left\{\hat{K}_W^R(t, s, \Delta^*) - \hat{K}_W^R(t, s, \Delta)\right\}. \tag{3.21}$$

The next step is to estimate the covariance operators K^V and K^W. Again, the method of moments estimator is used to get the estimate for the total covariance $K^R(t, s) = \mathrm{Cov}\left(R_{ij}(t), R_{ij}(s)\right)$,

$$\hat{K}^R(t, s) = \frac{\sum_i \sum_j \left\{R_{ij}(t) - \bar{R}_{i\cdot}(t)\right\} \left\{R_{ij}(s) - \bar{R}_{i\cdot}(s)\right\}}{\sum_i \sum_j 1} \tag{3.22}$$

and

$$\hat{K}^V(t,s) = \hat{K}^R(t,s) - \frac{1}{2|N(\Delta^*)|}\sum_{i,j}\sum_{j':|j'-j|\geq\Delta^*}\left\{R_{ij}(t) - R_{ij'}(t)\right\}\left\{R_{ij}(s) - R_{ij'}(s)\right\}. \quad (3.23)$$

By using equation (3.17), we can estimate $K^W(t,s)$ by

$$\hat{K}^W(t,s) = \frac{1}{2|N(\Delta^*)|}\sum_{i,j}\sum_{j':|j'-j|\geq\Delta^*}\left\{R_{ij}(t) - R_{ij'}(t)\right\}\left\{R_{ij}(s) - R_{ij'}(s)\right\} - \hat{\sigma}_S^2, \text{ for } t \neq s. \quad (3.24)$$

To avoid the nugget effect due to the white noise terms, $\hat{K}^W(t,s)$ for $t = s$ is obtained by performing thin-plate spline smoothing on $\hat{K}^W(t,s)$, $t \neq s$ obtained from the last equation. This approach was proposed by Staniswalis and Lee (1998).

We can apply spectral decomposition to obtain estimated eigenvalues and eigenfunctions for K^V and K^W. See Staicu et al. (2010) for details on various methods to select the numbers of components K_1 and K_2. Some examples of possible methods are choosing the numbers of the components such that the percent explained variance is higher than a specified threshold, or choosing them using a likelihood ratio test. The estimated principal component component scores are computed using best linear unbias prediction (BLUP). Plugging in $V_i(t)$ and $W_{ij}(t)$ in equation (3.13) to equation (3.9), we obtain a linear mixed model with no fixed effects, where the random effects are the principal component scores for V_i and W_{ij} and $\left\{S_{ij}\right\}$. These random effects can be estimated using Henderson's mixed model equation (Henderson, 1950). For derivation and discussion of BLUP in linear mixed model, see Robinson (1991). Finally, we estimate σ_ϵ^2 by

$$\hat{\sigma}_\epsilon^2 = \int\left\{\hat{L}^W(t,t) - \hat{K}^W(t,t) - \hat{v}^S(0)\right\}dt \quad (3.25)$$

where $\hat{L}^W(t,t)$ is $\hat{K}_W^R(t,s,\Delta^*)$ computed at $t = s$.

3.3 Longitudinal DTI Results

This section presents result from an analysis of parallel diffusivity (ParD) from 162 MS patients observed across 93 tract locations t, using the methodology presented in the previous section. The tract distance is normalized to [0,1]. The visit time j are in the unit of three-month intervals counting from the first visit. Here, we have data from $j = 0$ to $j = 17$. Table 3.1 shows the number of observations for each j. We assumed skew t distribution and used function `st.mle` in `R` package `sn` to obtain the maximum likelihood estimates as explained in Section 3.2.1. The estimated degree of freedom is 5.96. The undersmoothed estimates $\widetilde{\sigma}(t)$ and $\widetilde{\alpha}(t)$ for the standard deviation and shape parameter functions are smoothed further using function `gam` in `R` package `mgcv`. The smoothing parameters are selected using GCV. For the mean function estimates $\widetilde{\mu}_j(t)$, we fitted a tensor product spline using function `gam` in `R` package `mgcv` instead of a thin-plate spline, as a thin-plate spline smooths the same amount in both directions which is not appropriate in this application. We placed knots at the 9 visit times j that have the highest number of observations. Table 3.1 displays the number of observations for each j. We used 15 equally spaced knots for t.

Table 3.1: The number of observations for each visit time j

j	0	1	2	3	4	5	6	7	8
number of observations	162	14	44	10	73	18	12	5	25
j	9	10	11	12	13	14	15	16	17
number of observations	16	4	4	9	12	4	3	3	3

Figure 3.1 displays the final estimates for the mean functions after tensor product spline smoothing. The mean functions fluctuate the most at locations

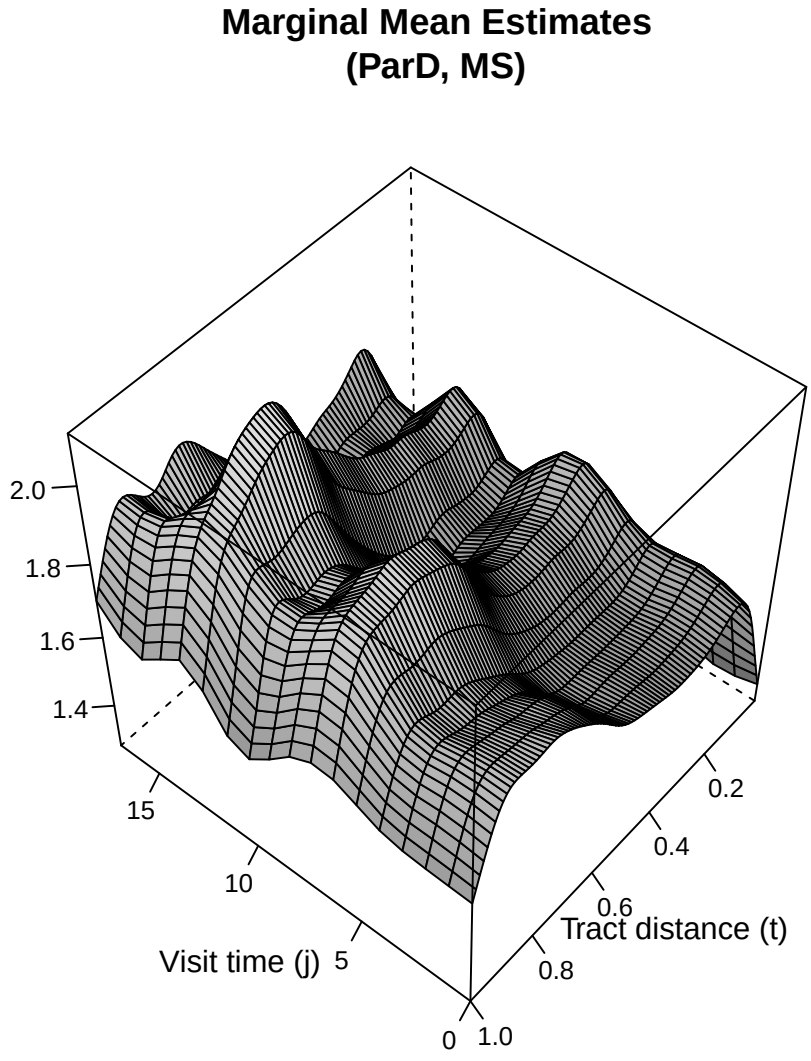

Figure 3.1: Estimated marginal mean functions.

0.6 to 0.9, and seem to show a little bit of trend to increase over time. Note that there are small number of observations beyond $j = 14$. Figure 3.2 shows the final estimates for the variance and skewness (which depends on both the shape parameters and the degree of freedom) functions. The estimated variance function is almost identical to the one in Figure 2.1 in the previous section, which includes only the data from first visit. The estimated skewness function is similar to the one in Figure 2.1 except for locations 0.2 to 0.4. The skewness in this

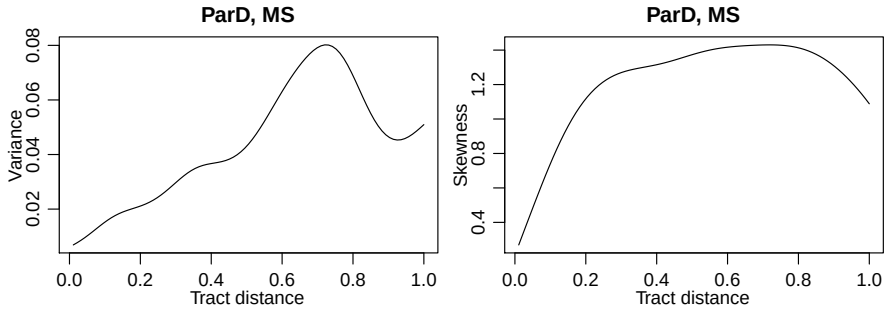

Figure 3.2: Estimated marginal variance and skewness functions.

range seems to be slightly higher here.

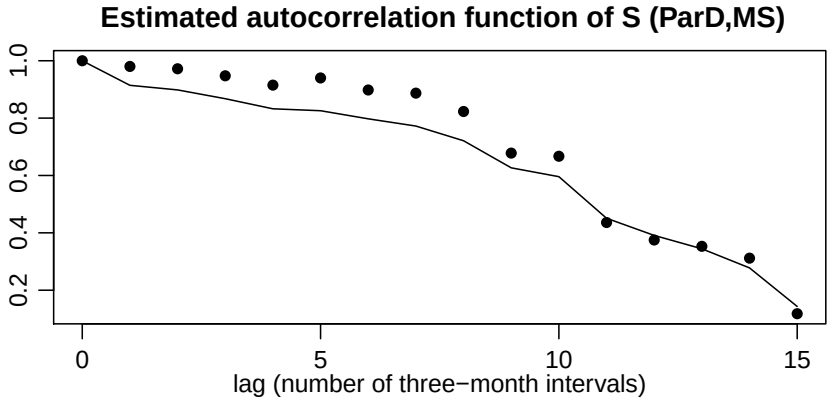

Figure 3.3: Estimated autocorrelation function of S .

The data was transformed using equation (3.1) and decomposed as in equation (3.9). Figure 3.3 shows the estimated autocorrelation of S_i in black dots. The black line shows the estimated autocorrelation function after positive-semidefinite adjustment. The adjustment was done using the `R` function `nearPD` in package `Matrix` (Bates and Maechler, 2013), which implements an algorithm based on an alternating projection method introduced by Higham (2002). The estimated variance of S_i, $\widehat{\nu}^S(0)$ is 0.22. The threshold Δ^* we selected from using crossvalidation is 16 and the width w is 1. The autocorrelation function significantly drops at lag 11 but doesn't drop below 0.2 until lag 15. Since Δ^*

is very close to 17 which is the largest lag we have in our data, we recommend that this step should be carefully investigated when the observed data is available for larger range of time. The residual plots we present later in this chapter suggest that our result is adequate for this data.

Table 3.2: Estimated eigenvalues at subject-specific and visit-specific levels

	subject-specific (V)						visit-specific (W)		
comp	1	2	3	4	5	6	1	2	3
eigenvalues	10.43	1.27	1.03	0.59	0.47	0.32	25.41	6.08	3.20
% variation	64.7	7.9	6.4	3.7	2.9	2.0	71.5	17.1	9.0
cumulative % variation	64.7	72.5	78.9	82.6	85.5	87.5	71.5	88.6	97.5

The likelihood ratio test selected $K_1 = 17$ and $K_2 = 8$. Table 3.2 provides some of the estimated eigenvalues for V (subject level) and W (visit level). These values indicate that the variability at the visit-level is about 2.5 times of the variability at the subject level.

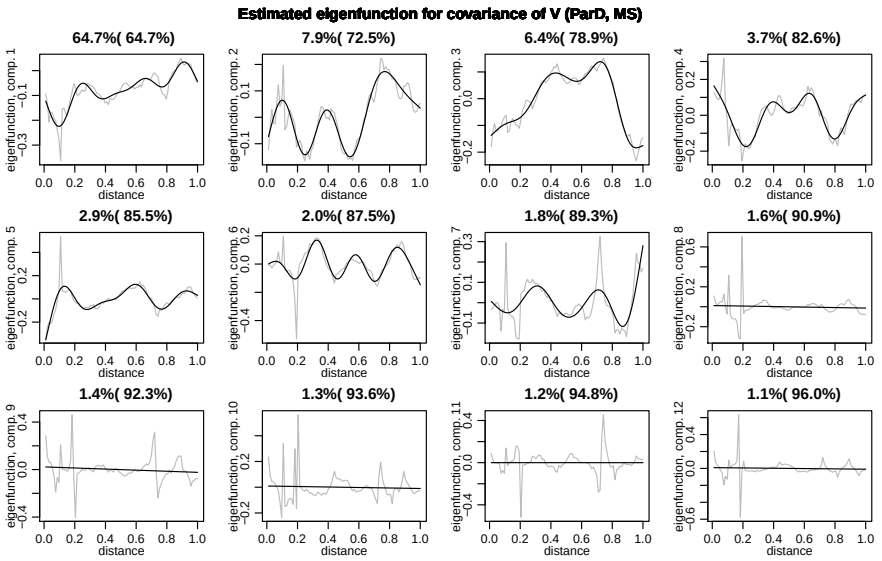

Figure 3.4: Estimated eigenfunctions of V .

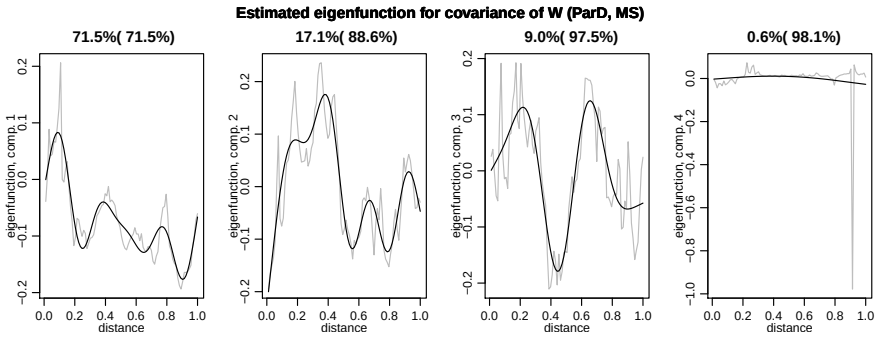

Figure 3.5: Estimated eigenfunctions of W .

At the subject-specific level, the first component explains about 65% of the variability. The first 12 components explain about 96% of the variability. At the visit-specific level, the first component explains about 71% of the variability and the first 3 components explain over 97% of the variability. Figures 3.4 and 3.5 display the first 12 and 4 eigenfunctions for V and W, respectively. The gray lines are the estimated eigenfunctions from the spectral decomposition. The black lines are the estimated eigenfunctions after penalized spline fitting using function `gam` in `R` package `mgcv`. None of the components 6-12 of V explains a lot of the variability, and their corresponding estimated eigenfunctions are close to a zero horizontal straight line. This is probably due to the variability of V is more difficult to capture than that of W in this data set.

Figure 3.6 shows the boxplots for the estimated functional principal component scores of V and W (for $j = 0$), respectively. Figure 3.7 displays the plots of covariance of V and W. The variance of V is highest at locations close to 0.1, while the variance of W is highest at locations close to 0.9.

Figure 3.8 shows a plot of sample correlation of all the residuals, where the

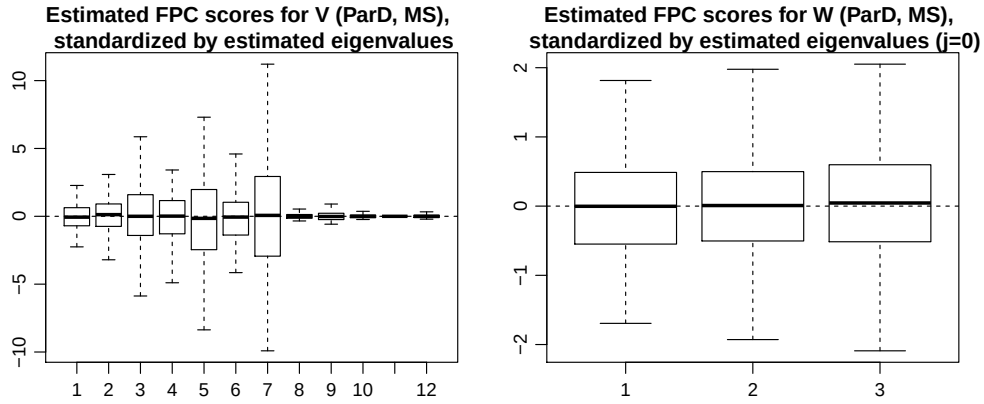

Figure 3.6: Estimated functional principal component scores of V and W ($j = 0$) .

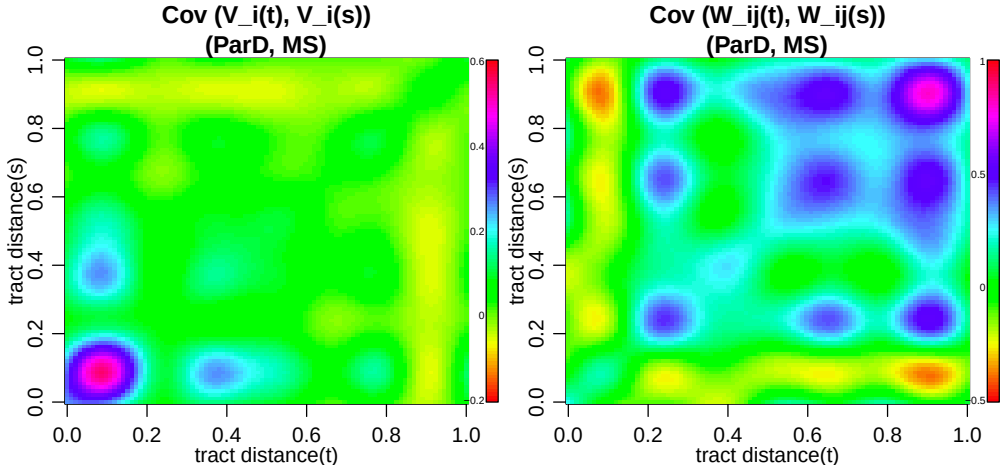

Figure 3.7: Estimated covariances of V and W .

residuals $\widehat{\epsilon_{ij}}$ are computed as

$$\widehat{\epsilon_{ij}} = R_{ij}(t) - \sum_{k_1=1}^{K_1} \hat{\xi}_{i,k_1} \hat{\phi}_{k_1}^{(1)}(t) - \sum_{k_2=1}^{K_2} \hat{\zeta}_{ij,k_2} \hat{\phi}_{k_2}^{(2)}(t) - \hat{S}_{ij} \tag{3.26}$$

where all the random components are estimated using BLUP. The corrrelation seems to be close to 0 in most region except for the part close to the diagonal, indicating that our model captures the correlation between tract locations that are not too close quite well, but not between the locations that are less than

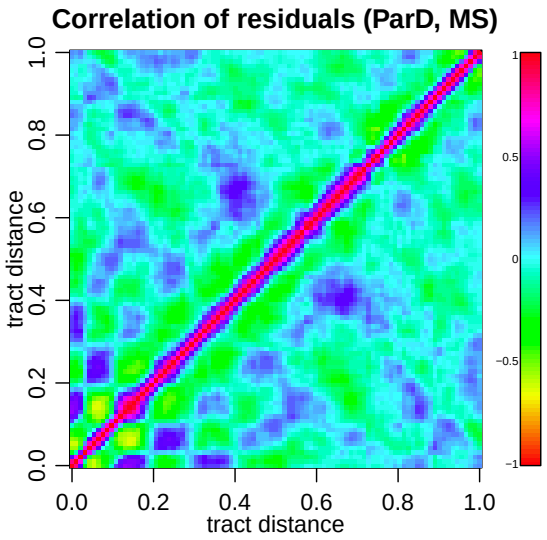

Figure 3.8: Correlation of all residuals.

about 0.1 units apart. One possible remedy for this problem is to smooth the transformed processes $R_{ij}(t)$ before decomposing them.

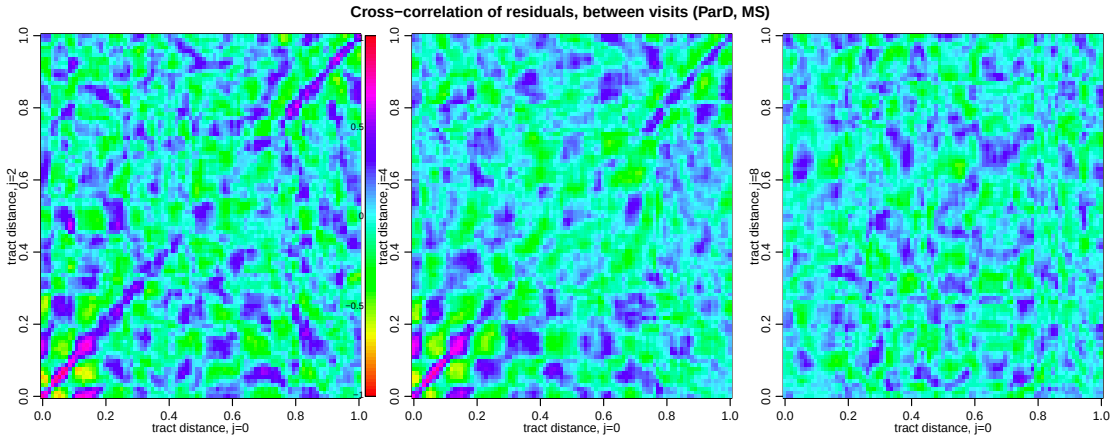

Figure 3.9: Cross-correlation of residuals from different visits.

We look at the residuals in more details in Figure 3.9 which shows the cross-correlation between residuals from different j. The first plot shows the cross-correlation between $j = 0$ and $j = 2$. The plot shows high correlation in small areas at both ends of the diagonal, indicating a problem of capturing correlation

between locations that are close together at the two ends of the tract. The pattern is less apparent in the second plot for the cross-correlation of residuals from $j = 0$ and $j = 4$. There doesn't seem to be any pattern in the last plot which shows the cross-correlation of residuals from $j = 0$ and $j = 8$. The first plot in Figure 3.9 suggests that it might be useful to split the tract to 3 segments, left, middle and right, and allow $W_{ij}(t)$ to have different covariance structure for t in different segments. This would increase the complexity of the model and is possibly doable when the data set is fairly large.

3.4 Discussion

We found that the mean of parallel diffusivity in MS patients tend to increase over time. There is not much change in the variance and skewness estimates when we included data from multiple visits. From our copula modeling, we found that the autocorrelation doesn't drop below 0.2 until lag 15, suggesting the persistence of autocorrelation over long period of time of the parallel diffusivity. The variance of the second-level visit-specific that depends on the tract location $W_{ij}(t)$ is about 2.5 times of that of the first level subject-specific $V_i(t)$. This indicate that the variation across visits is higher than the variation across subjects, and the latter can possibly be more difficult to model in this DTI application.

We found that our model has problems capturing the correlation between the tract locations that are very close together. One possible solution is to smooth the transformed process before proceeding with the copula modeling. Also, as a larger data set becomes available, it might be useful to split the tract

into segments and allow the covariance structure to be different in different segments. Sufficiently large data set would also allow the degree of freedom parameter in the skew-t marginal distribution to vary across the tract locations, and would possibly lead to better understanding of the change of DTI imaging along the tract.

APPENDIX A

COMPUTATION FOR EM ALGORITHM

This appendix explains the detail for the EM algorithm used to estimate parameters $\sigma^2_{\epsilon p}$, Θ_p and Σ.

A.1 The E-Step

Let $E^*(\cdot)$ denote the expectation given the observed data $\{R_{ip}\}$. The goal of this step is to compute $E^*(Z_{ip})$, $E^*(Z_{ip}Z_{ip}^T)$ and $E^*(Z_{ip}, Z_{ip'})$. Suppose $Z_i|R_i \sim \mathrm{N}(m_i, V_i)$. That is,

$$f(Z_i|R_i) \propto \exp\left\{-(1/2)(Z_i - m_i)^T V_i^{-1}(Z_i - m_i)\right\}. \tag{A.1}$$

Note that $f(Z_i|R_i) \propto f(Z_i, R_i)$ and

$$\begin{aligned} f(Z_i, R_i) &\propto \exp\left[-\frac{1}{2}\left\{\sum_p \frac{1}{\sigma^2_{\epsilon p}}\left(R_{ip} - B\Theta_p Z_{ip}\right)^T\left(R_{ip} - B\Theta_p Z_{ip}\right) + Z_i^T\Sigma^{-1}Z_i\right\}\right] \\ &\propto \exp\left\{-\frac{1}{2}\left(\sum_p \frac{1}{\sigma^2_{\epsilon p}} Z_{ip}^T\Theta_p^T B^T B\Theta_p Z_{ip} + Z_i^T\Sigma^{-1}Z_i\right) + \sum_p \frac{1}{\sigma^2_{\epsilon p}} R_{ip}^T B\Theta_p Z_{ip}\right\} \\ &= \exp\left\{-\frac{1}{2}Z_i^T(E^TE + \Sigma^{-1})Z_i + M_i Z_i\right\}, \end{aligned} \tag{A.2}$$

where

$$M_i = \left[\begin{array}{cccc} \frac{1}{\sigma^2_{\epsilon 1}} R_{i1}^T B\Theta_1 & \frac{1}{\sigma^2_{\epsilon 2}} R_{i2}^T B\Theta_2 & \cdots & \frac{1}{\sigma^2_{\epsilon P}} R_{iP}^T B\Theta_P \end{array}\right]_{1\times\sum_p K_p}, \tag{A.3}$$

$$E = \left[\begin{array}{ccc} \frac{1}{\sigma_{\epsilon 1}} B\Theta_1 & & 0 \\ & \ddots & \\ 0 & & \frac{1}{\sigma_{\epsilon P}} B\Theta_P \end{array}\right]_{mP\times\sum_p K_p} \tag{A.4}$$

and Σ is the covariance matrix of Z_i as defined in equation (1.9).

Comparing the two expressions, we have that

$$V_i = \left(E^T E + \Sigma^{-1}\right)^{-1} = V; \tag{A.5}$$

$$m_i = V^T M_i^T = V M_i^T. \tag{A.6}$$

Decompose the matrix V and vectors m_i obtained from above as

$$V = \begin{pmatrix} V_{11} & \cdots & V_{1P} \\ \vdots & \ddots & \vdots \\ V_{P1} & \cdots & V_{PP} \end{pmatrix}, m_i = \begin{pmatrix} m_{i1} \\ \vdots \\ m_{iP} \end{pmatrix}. \tag{A.7}$$

The dimension of $V_{pp'}$ and m_{ip} correspond to the dimension of Z_{ip}. The predictions for Z_{ip} and their moments required by the EM algorithm are

$$\begin{aligned} E^*(Z_{ip}) &= m_{ip}; \\ E^*(Z_{ip} Z_{ip}^T) &= m_{ip} m_{ip}^T + V_{pp}; \\ E^*(Z_{ip} Z_{ip'}^T) &= m_{ip} m_{ip'}^T + V_{pp'}. \end{aligned} \tag{A.8}$$

A.2 The M-Step

In the optimization, parameters $\sigma_{\epsilon p}^2$, Θ_p and Σ are updated in order. The values of the parameters on the right hand side of the last equation in steps 1,2 and 3 are plugged in using their current values. The details are as follows.

Step 1: Update the estimates for $\sigma_{\epsilon p}^2$, $p = 1, \ldots, P$. We want to minimize

$$E^* \left[\sum_{i=1}^{N} \left\{ m \log(\sigma_{\epsilon p}^2) + \frac{1}{\sigma_{\epsilon p}^2} \left(R_{ip} - B\Theta_p Z_{ip}\right)^T \left(R_{ip} - B\Theta_p Z_{ip}\right) \right\} \right]. \tag{A.9}$$

The expectation of the quadratic form can be written as

$$\begin{aligned}
& E^*\left\{\left(R_{ip} - B\Theta_p Z_{ip}\right)^T \left(R_{ip} - B\Theta_p Z_{ip}\right)\right\} \\
= \; & R_{ip}^T R_{ip} - 2R_{ip}^T B\Theta_p E^*(Z_{ip}) + E^*\left(Z_{ip}^T \Theta_p^T B^T B\Theta_p Z_{ip}\right) \\
= \; & R_{ip}^T R_{ip} - 2R_{ip}^T B\Theta_p m_{ip} + \mathrm{tr}\left\{(\Theta_p^T B^T B\Theta_p) E^*(Z_{ip} Z_{ip}^T)\right\} \\
= \; & R_{ip}^T R_{ip} - 2R_{ip}^T B\Theta_p m_{ip} + \mathrm{tr}\left\{(\Theta_p^T B^T B\Theta_p)(V_{pp} + m_{ip} m_{ip}^T)\right\} \\
= \; & R_{ip}^T R_{ip} - 2R_{ip}^T B\Theta_p m_{ip} + \mathrm{tr}(m_{ip}^T \Theta_p^T B^T B\Theta_p m_{ip}) + \mathrm{tr}(B\Theta_p V_{pp} \Theta_p^T B^T) \\
= \; & (R_{ip} - B\Theta_p m_{ip})^T (R_{ip} - B\Theta_p m_{ip}) + \mathrm{tr}(B\Theta_p V_{pp} \Theta_p^T B^T). \qquad \text{(A.10)}
\end{aligned}$$

So, we want to minimize

$$Nm \log(\sigma_{\epsilon p}^2) + \frac{1}{\sigma_{\epsilon p}^2} \sum_{i=1}^{N} \left\{(R_{ip} - B\Theta_p m_{ip})^T (R_{ip} - B\Theta_p m_{ip}) + \mathrm{tr}(B\Theta_p V_{pp} \Theta_p^T B^T)\right\}. \quad \text{(A.11)}$$

Hence, we update $\sigma_{\epsilon p}^2$ by

$$\widehat{\sigma}_{\epsilon p}^2 = \frac{1}{Nm} \sum_{i=1}^{N} \left\{(R_{ip} - B\Theta_p m_{ip})^T (R_{ip} - B\Theta_p m_{ip}) + \mathrm{tr}(B\Theta_p V_{pp} \Theta_p^T B^T)\right\}, \qquad \text{(A.12)}$$

for each $p = 1, \ldots, P$.

Step 2: Update Θ_p, by updating each column Θ_{pk} sequentially. For each k, we want to minimize

$$\sum_{i=1}^{N} E^*\left(\|R_{ip} - \sum_{l \neq k} B\Theta_{pl} Z_{ipl} - B\Theta_{pk} Z_{ipk}\|^2\right) + \sigma_{\epsilon p}^2 \lambda_p \Theta_{pk}^T \int b''(t) b''(t)^T dt \Theta_{pk}. \quad \text{(A.13)}$$

Similar to the first step, the expression is minimized when

$$\begin{aligned}
\widehat{\Theta}_{pk} \; = \; & \left\{\sum_{i=1}^{N} E^*(Z_{ipk}^2) B^T B + \sigma_{\epsilon p}^2 \lambda_p \int b''(t) b''(t)^T dt\right\}^{-1} \\
& \times \sum_{i=1}^{N} B^T \left\{R_{ip} E^*(Z_{ipk}) - \sum_{l \neq k} B\Theta_{pl} E^*(Z_{ipk} Z_{ipl})\right\}, \qquad \text{(A.14)}
\end{aligned}$$

where $E^*(Z_{ipk}^2)$, $E^*(Z_{ipk})$ and $E^*(Z_{ipk} Z_{ipl})$ are the $(k,k)-, k-$ and $(k,l)-$ element of $m_{ip} m_{ip}^T + V_{pp}$, m_{ip} and $m_{ip} m_{ip'}^T + V_{pp'}$, respectively.

Step 3: Update Σ. We want to minimize

$$E^*\left\{\sum_{i=1}^{N}\left(\log|\Sigma| + Z_i^T\Sigma^{-1}Z_i\right)\right\} = \sum_{i=1}^{N}\left[\log|\Sigma| + \mathrm{tr}\left\{\Sigma^{-1}(V + m_i m_i^T)\right\}\right]. \tag{A.15}$$

Since $S = \sum_i \left(V + m_i m_i^T\right)$ is positive-definite, there is a unique positive-definite matrix $S^{1/2}$ such that $S^{1/2}S^{1/2} = S$. Let $W = S^{1/2}\Sigma^{-1}S^{1/2}$. Then we want to minimize

$$N\log|\Sigma| + \mathrm{tr}(S\Sigma^{-1}) = -N\log|W| + N\log|S| + \mathrm{tr}(W). \tag{A.16}$$

The positive-definite W can be diagonalized to get the diagonal elements $\tau_1, \ldots, \tau_K$, where $K = \sum_{p=1}^{P} K_p$. What we finally want to minimize is $-N\sum_{k=1}^{K}\log\tau_k + \sum_{k=1}^{K}\tau_k$. The solution is $\tau_k = N$ for all k. That is, $W = NI_K$ which implies that $\Sigma = S^{1/2}W^{-1}S^{1/2} = \frac{1}{N}S$. Hence, in this step, we update Σ by

$$\widehat{\Sigma} = \frac{1}{N}S = \frac{1}{N}\sum_i (V + m_i m_i^T). \tag{A.17}$$

Step 4: Orthogonalization and update of the covariance matrix. Suppose the matrix $\widehat{\Sigma}$ obtained in step 3 takes the form

$$\widehat{\Sigma} = \begin{pmatrix} \Lambda_{11} & \Lambda_{12} & \cdots & \Lambda_{1P} \\ \Lambda_{21} & \Lambda_{22} & \cdots & \Lambda_{2P} \\ \vdots & \vdots & \ddots & \vdots \\ \Lambda_{P1} & \Lambda_{P2} & \cdots & \Lambda_{PP} \end{pmatrix}. \tag{A.18}$$

Let $\widehat{\Theta}_p\Lambda_{pp}\widehat{\Theta}_p^T = Q_p S_p Q_p^T$ be the eigenvalue decomposition in which Q_p has orthogonal columns, and S_p is diagonal with diagonal elements in a decreasing order. We update $\widehat{\Theta}_p$ by Q_p and Λ_{pp} by S_p. Since the updating in this step corresponds to the transformation $Z_{ip} \leftarrow Q_p^T\widehat{\Theta}_p Z_{ip}$, we update $\widehat{C}_{pp'}$ by $Q_p^T\widehat{\Theta}_p\Lambda_{pp'}\widehat{\Theta}_{p'}^T Q_{p'}$. The sign of each column of Θ is changed as necessary so that the elements with the largest magnitude of each column are positive.

BIBLIOGRAPHY

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics* **12,** 171–178.

Azzalini, A. and Capitanio, A. (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew *t* distribution. *Journal of the Royal Statistics Society, Series B* **65,** 367–389.

Azzalini, A. (2011). *R Package 'sn': The skew-normal and skew-t distributions (version 0.4-17)*. URL `http://azzalini.stat.unipd.it/SN`, accessed August 1, 2012.

Bates, D., and Maechler M. (2013). *R Package 'Matrix': Sparse and Dense Matrix Classes and Methods (version 1.0-12)*. URL `http://cran.r-project.org/web/packages/Matrix/index.html`, accessed July 15, 2013.

Basser, P. J., Mattiello, J., and LeBihan, D. (1994). MR diffusion tensor spectroscopy and imaging. *Biophysical Journal* **66,** 259–267.

Basser, P. J., Pajevic, S., Pierpaoli, C., Duda, J., and Aldroubi, A. (2000). In vivo fiber tractography using dt-mri data. *Magnetic Resonance in Medicine* **44,** 625–632.

Crainiceanu, C.M., Reiss, P., Goldsmith, J., Huang, L., Huo, L., Scheipl, F. et al. (2012). *R Package 'refund': Regression with functional data (version 0.1-6)*. URL `http://cran.r-project.org/web/packages/refund/index.html`.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B* **39,** 1–38.

Di C., Crainiceanu C.M., Caffo B.S., and Punjabi N.M. (2009). Multilevel Functional Principal Component Analysis. *The Annals of Applied Statistics* **3,** 458-488.

Duchon, J. (1975). *Fonctions splines et vecteurs aleatoires,* Tech. Report 213, Seminaire d' Analyse Numerique, Universite Scientifique et Medicale, Grenoble.

Duchon, J. (1976). *Fonctions-splines et esperances conditionnelles de champs gaussiens,* Ann. Sci. Univ. Clermont Ferrand II Math., 14, 19-27 Report 213, Seminaire d' Analyse Numerique, Universite Scientifique et Medicale, Grenoble.

Duchon, J. (1977). *Splines minimizing rotation-invariant semi-norms in Sobolev spaces,* in Constructive Theory of Functions of Several Variables, Springer-Verlag, Berlin.

Ferraty, F. and Romain, Y. (2010). *The Oxford Handbook of Functional Data Analysis.* New York: Oxford University Press.

Furrer, R., Nychka D., and Sain S. (2013). *R Package 'fields': Tools for spatial data (version 6.8)*. URL `http://cran.r-project.org/web/packages/fields/index.html`, accessed July 20, 2013.

Green, P.J. and Silverman, B.W. (2010). *Nonparametric Regression and Generalized Linear Models.* London: Chapman and Hall.

Greven, S., Crainiceanu, C. M., Caffo, B., and Reich, D. S. (2010). Longitudinal functional principal component analysis. *Electronic Journal of Statistics* **4,** 1022–1054.

Goldsmith, J., Feder, J., Crainiceanu, C. M., Caffo, B., and Reich, D. S. (2011a).

Penalized functional regression. *Journal of Computational and Graphical Statistics* **20,** 830–851.

Goldsmith, J., Crainiceanu, C. M., Caffo, B., and Reich, D. S. (2011b). Penalized functional regression analysis of white-matter tract profiles in multiple sclerosis. *NeuroImage* **57,** 431–439.

Goldsmith, J., Crainiceanu, C. M., Caffo, B., and Reich, D. S. (2012). Longitudinal penalized functional regression for cognitive outcomes on neuronal tract measurements. *Journal of the Royal Statistical Society, Series C* **61,** 453–469.

Henderson, C. R.(1950). Estimation of genetic parameters (abstract). *Annals of Mathematical Statistics* **21,** 309–310.

Higham, N. (2002). Computing the nearest correlation matrix - a problem from finance. *IMA Journal of Numerical Analysis* **22,** 329–343.

Levy, B. C. (2008). *Principles of Signal Detection and Parameter Estimation.* New York: Springer.

Li, Y. and Ruppert, D. (2008). On the asymptotics of penalized splines. *Biometrika* **95,** 415–436.

McLean, M. W., Hooker, G., Staicu, A., Scheipl, F., and Ruppert, D. (2012). Functional generalized additive models. *Journal of Computational and Graphical Statistics,* to appear.

Meinguet, J. (1979). *Multivariate interpolation at arbitrary points made simple,* J. Appl. Math. Phys. (ZAMP), 30, 292-304.

McNeil, A., Frey, R., and Embrechts, P. (2005). *Quantitative Risk Management.* Princeton: Princeton University Press.

Ozturk, A., Smith, S.A., Gordon-Lipkin E.M., Harrison, D.M., Shiee, N., Pham, D.L. et al. (2010). MRI of the corpus callosum in multiple sclerosis: association with disability. *Multiple Sclerosis* **16,** 166–177.

Ramsay, J. and Silverman, B. W. (2005). *Functional Data Analysis.* New York: Springer.

Redd, A. (2011). *R Package 'orthogonalsplinebasis': Orthogonal bspline basis functions (version 0.1.5)*. URL `http://osplinebasis.r-forge.r-project.org`, accessed February 1, 2013.

Reich, D. S., Smith, S. A., Zackowski, K. M., Gordon-Lipkin, E. M., Jones, C. K., Farrel, J. A.D. et al. (2005). Multiparametric magnetic resonance imaging analysis of the corticospinal tract in multiple sclerosis. *NeuroImage* **38,** 271–279.

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J. C. et al. (2011). pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* **12,** 77–84.

Robinson, G. K. (1991). That BLUP is a good thing: the estimation of random effects. *Statistical Science* **6,** 15–51.

Ruppert, D. (2002). Selecting the number of knots for penalized splines. *Journal of Computational and Graphical Statistics* **11,** 735–757.

Ruppert, D., Wand, M. P., and Carroll, R. J. (2003). *Semiparametric Regression,* Cambridge: Cambridge University Press.

Staicu, A.-M., Crainiceanu, C. M., and Carroll, R. J. (2010). Fast methods for spatially correlated multilevel functional data. *Biostatistics* **11,** 177-194.

Staicu, A., Crainiceanu, C. M., Reich, D. S., and Ruppert, D. (2011). Modeling functional data with spatially heterogeneous shape characteristics. *Biometrics* **68,** 331–343.

Staniswalis, J. G., and Lee, J. J. (1998). Nonparametric regression analysis of longitudinal data. *Journal of the American Statistical Association* **93,** 1403-1418.

Wahba, G. (1990). *Spline models for observational data.* CBMS-NSF Reg. Conf. Ser. Appl. Math. **59**.

Wood, S. N. (2003). Thin plate regression splines. *Journal of the Royal Statistics Society, Series B* **65,** 95–114.

Wood, S. N. (2006). *Generalized Additive Models: An Introduction with R.* Boca Raton, FL: Chapman & Hall/CRC.

Zhou, L., Huang, J. Z., and Carroll, R. J. (2008). Joint modelling of paired sparse functional data using principal components. *Biometrika* **95,** 601–619.